

From: Fredman, Jeffrey
Sent: Wednesday, January 19, 2005 10:34 AM
To: Schreiber, David
Cc: Chunduru, Suryaprabha
Subject: FW: ref to rush sequence search for case # 09/603,665

David,

I Approve the Rush.

I am not exactly sure what Prabha means to search, but perhaps she means the 20-30 nucleotides overlapping each of the listed positions.

Jeff Fredman

-----Original Message-----

From: Chunduru, Suryaprabha
Sent: Wednesday, January 19, 2005 9:41 AM
To: Fredman, Jeffrey
Subject: ref to rush sequence search for case # 09/603,665

Hi Jeff,

I request you to approve a rush sequence search for the above amended case.
Oligomer search for sequences comprising positions 1694, 1854, 1967, 2017, 2050 of SEQ ID NO. 5.
I request David Schreiber to perform the search.

thanks
Suryaprabha Chunduru
AU 1637, Room No. 2C29,
Mail box 2C18
tel. # 272-0783.

This Page Blank (uspto)

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: <u>D. Schweizer</u>	NA Sequence (#) _____	STN _____
Searcher Phone #: <u>272-2526</u>	AA Sequence (#) <u>9</u>	Dialog _____
Searcher Location: <u>Rensen E01A61</u>	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: _____	Bibliographic _____	Dr. Link _____
Date Completed: <u>2/2</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep. Review Time <u>15</u>	Fulltext _____	Sequence Systems <u>Campagne E</u>
Clerical Prep. Time: _____	Patent Family _____	WWW/Internet _____
Online Time <u>61</u>	Other _____	Other (specify) _____

This Page Blank (uspto)



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 144006

TO: Suryaprabha Chunduru
Location: REM/2C29/2C18
Art Unit: 1637
Wednesday, February 02, 2005

Case Serial Number: 09/603665

From: David Schreiber
Location: Biotech-Chem Library
Remsen E01A61
Phone: 272-2526

david.schreiber@uspto.gov

Search Notes

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:58:28 ; Search time 17.5556 Seconds
(Without alignments)
154.882 Million cell updates/sec

Title: SEQ5ASN1694

Perfect score: 41 LKLLCNFGAENPDFPVVL.....TAVKLIAPEKKEKNVLSA 41

Sequence: 1 LKLLCNFGAENPDFPVVL.....TAVKLIAPEKKEKNVLSA 41

Scoring table: OLIGO Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size : 0

Total number of hits satisfying chosen parameters: 290433

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents AA:*

- 1: /cgn2_6/prodata/1/1aa/5A COMB.pep.*
- 2: /cgn2_6/prodata/1/1aa/5B COMB.pep.*
- 3: /cgn2_6/prodata/1/1aa/6A COMB.pep.*
- 4: /cgn2_6/prodata/1/1aa/6B COMB.pep.*
- 5: /cgn2_6/prodata/1/1aa/PCITUS COMB.pep.*
- 6: /cgn2_6/prodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	17.1	160	4 US-09-270-767-33214	Sequence 33214, A
2	7	17.1	160	4 US-09-270-767-48431	Sequence 48431, A
3	6	14.6	51	4 US-09-513-999C-5558	Sequence 5558, Ap
4	6	14.6	61	4 US-09-583-110-4172	Sequence 4172, Ap
5	6	14.6	81	4 US-09-621-876-6456	Sequence 6456, Ap
6	6	14.6	87	4 US-09-513-999C-4616	Sequence 4616, Ap
7	6	14.6	88	4 US-09-248-796A-25228	Sequence 25228, A
8	6	14.6	98	4 US-09-540-236-3312	Sequence 3312, Ap
9	6	14.6	99	4 US-09-270-767-35660	Sequence 35660, A
10	6	14.6	99	4 US-09-270-767-38390	Sequence 38390, A
11	6	14.6	127	4 US-09-270-767-50877	Sequence 50877, A
12	6	14.6	127	4 US-09-270-767-53607	Sequence 53607, A
13	6	14.6	135	6 5229115-2	Sequence 22195, A
14	6	14.6	135	4 US-09-248-796A-22195	Sequence 22195, A
15	6	14.6	147	4 US-09-710-279-1214	Sequence 1214, Ap
16	6	14.6	154	1 US-08-466-908-2	Sequence 2, Appl
17	6	14.6	154	1 US-08-231-205A-2	Sequence 2, Appl
18	6	14.6	154	1 US-08-871-161-2	Sequence 2, Appl
19	6	14.6	159	4 US-09-248-796A-19969	Sequence 19969, A
20	6	14.6	159	3 US-08-737-248-23	Sequence 23, Appl
21	6	14.6	205	4 US-09-252-991A-26704	Sequence 26704, A
22	6	14.6	218	4 US-09-710-279-450	Sequence 450, App
23	6	14.6	226	4 US-09-248-796A-18711	Sequence 18711, A
24	6	14.6	237	4 US-09-248-796A-18639	Sequence 18639, A
25	6	14.6	253	4 US-09-248-796A-14685	Sequence 14685, A
26	6	14.6	253	2 US-08-685-992-10	Sequence 10, Appl
27	6	14.6	253	2 US-09-144-925-10	Sequence 10, Appl

28	6	14.6	294	4 US-09-540-236-3779	Sequence 3779, Ap
29	6	14.6	298	4 US-09-134-000C-5845	Sequence 5845, Ap
30	6	14.6	309	4 US-09-489-039A-8203	Sequence 8203, Ap
31	6	14.6	325	3 US-09-134-001C-3551	Sequence 3551, Ap
32	6	14.6	329	4 US-09-149-476-483	Sequence 483, App
33	6	14.6	339	4 US-09-107-532A-5514	Sequence 5514, Ap
34	6	14.6	344	4 US-09-248-796A-16383	Sequence 16383, A
35	6	14.6	348	1 US-08-454-196-8	Sequence 8, Appl
36	6	14.6	348	1 US-08-454-196-17	Sequence 17, Appl
37	6	14.6	348	3 US-09-064-033-17	Sequence 8, Appl
38	6	14.6	348	3 US-09-064-033-17	Sequence 17, Appl
39	6	14.6	348	4 US-09-291-046-8	Sequence 8, Appl
40	6	14.6	348	4 US-09-291-046-17	Sequence 17, Appl
41	6	14.6	351	4 US-09-248-796A-16176	Sequence 16176, A
42	6	14.6	372	2 US-08-683-2628-64	Sequence 64, Appl
43	6	14.6	372	3 US-09-161-707-64	Sequence 5, Appl
44	6	14.6	377	3 US-09-150-133-5	Sequence 5, Appl
45	6	14.6	377	3 US-09-150-141-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-09-270-767-33214
; Sequence 33214, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 33214
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-33214

Query Match 17.1%; Score 7; DB 4; Length 160;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 VPVLTNTA 23
Db 52 VPVLTNTA 58

RESULT 2
US-09-270-767-48431
; Sequence 48431, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 48431
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-48431

Query Match 17.1%; Score 7; DB 4; Length 160;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 17 VPVLTNTA 23

6mer

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:53:01 ; Search time 69.7778 Seconds
(without alignments)
210.782 Million cell updates/sec

Title: SE05GLY2017
Perfect score: 41
Sequence: 1 NCLYKIFLPTQHIFSKRA.....ALMPLVDQLNRLGSEK 41

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 2002273 seqs, 358729299 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1316031

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: A_Geneseq_23Sep04.*

1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	515	2	AAW54099
2	20	48.8	349	4	AAW54099
3	20	48.8	1149	7	AAW54099
4	20	48.8	2144	4	AAW54099
5	7	17.1	118	7	AAW54099
6	7	17.1	650	7	AAW54099
7	7	17.1	655	7	AAW54099
8	7	17.1	655	7	AAW54099
9	6	14.6	36	4	AAW54099
10	6	14.6	36	4	AAW54099
11	6	14.6	36	4	AAW54099
12	6	14.6	36	4	AAW54099
13	6	14.6	36	4	AAW54099
14	6	14.6	36	4	AAW54099
15	6	14.6	36	4	AAW54099
16	6	14.6	36	4	AAW54099
17	6	14.6	36	4	AAW54099
18	6	14.6	36	4	AAW54099
19	6	14.6	36	4	AAW54099
20	6	14.6	36	4	AAW54099
21	6	14.6	36	4	AAW54099
22	6	14.6	36	4	AAW54099
23	6	14.6	36	4	AAW54099
24	6	14.6	36	4	AAW54099
25	6	14.6	36	4	AAW54099

ALIGNMENTS

26	6	14.6	83	4	AAW55323	AAW55323 Human bra
27	6	14.6	83	4	ABG49358	ABG49358 Human liv
28	6	14.6	83	5	ABG37283	ABG37283 Human pep
29	6	14.6	83	5	ADK34692	ADK34692 Novel hum
30	6	14.6	86	4	AAO07992	AAO07992 Human pol
31	6	14.6	87	7	ADM26908	ADM26908 Hypertro
32	6	14.6	95	3	AAO1570	AAO1570 Human sec
33	6	14.6	102	5	AAO17456	AAO17456 Human liv
34	6	14.6	107	4	AB119971	AB119971 Human STR
35	6	14.6	108	4	ABG13882	ABG13882 Novel hum
36	6	14.6	138	3	AAW42068	AAW42068 Human ORF
37	6	14.6	138	4	AAO10626	AAO10626 Human pol
38	6	14.6	141	2	AAV07239	AAV07239 Fragment
39	6	14.6	141	2	AAV07241	AAV07241 Fragment
40	6	14.6	141	2	AAV07240	AAV07240 Fragment
41	6	14.6	141	5	ADK36505	ADK36505 Novel hum
42	6	14.6	150	4	ADM19866	ADM19866 Protein e
43	6	14.6	152	4	ABG64818	ABG64818 Drosophila
44	6	14.6	161	2	AAW77507	AAW77507 Human neu
45	6	14.6	164	4	AAW80897	AAW80897 Human hae

RESULT 1
ID AAW54099 standard; protein: 515 AA.

AC AAW54099;

DT 28-SEP-1998 (first entry)

DE Homo sapiens BAP28 sequence.

KW BARD1, ring protein; BRCA1, breast cancer; risk; diagnosis.

OS Homo sapiens.

PN MO9812327-A2.

PD 26-MAR-1998.

PF 19-SEP-1997; 97WO-US016842.

PR 20-SEP-1996; 96US-0025296P.

PR 03-APR-1997; 97US-0042611P.

PR 04-APR-1997; 97US-0042985P.

PA (TEXA) UNIV TEXAS SYSTEM.

PI Bowcock AM, Baer R;

DR WPI; 1998-230317/20.

DR N-PSDB; AAV24135.

XX DNA sequence encoding BARD1, B123, BE2, B14, B131 or B1445 - which as

XX breast cancer antigen, BRCA1, binding proteins are useful to identify

XX patient having or at risk of developing cancer.

XX Disclosure; Page 287-288; 348pp; English.

XX The sequence is that of a protein which can be used in the preparation of

XX the recombinant breast cancer antigen, BRCA1, binding proteins BARD1,

XX B123, BE2, B14, B131 or B1445, or a composition for the detection of a

XX wild type BARD1 composition for the detection or purification of BRCA1,

XX useful to identify a patient having, or at risk of developing cancer.

XX BARD1 can be used in the preparation of an anti-BARD1 antibody, and in

XX the detection and purification of a BRCA1 protein. BARD1, B123, BE2,

XX B14, B131 or B1445 can be used in the identification of a binding protein

XX agonist or antagonist that alters the binding of BARD1, B123, BE2, B14,

XX B131 or B1445 to BRCA1 or the biological activity of the BRCA1-BARD1,

This Page Blank (uspto)

CC B123, BE2, BE14, BE31 or BE445 complex. The antibodies can be used to
 CC detect BARD1, B123, BE2, BE14, BE31 or BE445, a specific anti-BARD1
 CC antibody can be used to identify a patient having or at risk of
 CC developing cancer

XX
 SQ Sequence 515 AA;

Query Match 100.0%; Score 41; DB 2; Length 515;
 Best Local Similarity 100.0%; Pred. No. 5 6e-37;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NCLYKIFLFDTOHFISKERAGALMPLVDQLENRLGGEK 41
 |||||
 DB 368 NCLYKIFLFDTOHFISKERAGALMPLVDQLENRLGGEK 408

RESULT 2
 AAB92729
 ID AAB92729 standard; protein; 349 AA.

XX
 AC AAB92729;

XX
 DT 26-JUN-2001 (first entry)

XX
 DE Human protein sequence SEQ ID NO:11159.

XX
 KM Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX
 OS Homo sapiens.

XX
 PN BP1074617-A2.

XX
 PD 07-FEB-2001.

XX
 PF 28-JUL-2000; 2000EP-00116126.

XX
 PR 29-JUL-1999; 99JP-00248036.

XX
 PR 27-AUG-1999; 99JP-00300253.

XX
 PR 11-JAN-2000; 2000JP-00118776.

XX
 PR 02-MAY-2000; 2000JP-00183767.

XX
 PR 09-JUN-2000; 2000JP-00241899.

XX
 PA (HELI-) HELIX RES INST.

XX
 PI Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX
 PI Iehli S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX
 PI WPI; 2001-318749/34.

XX
 PS Claim 8; SEQ ID NO 11159; 2537pp + Sequence Listing; English.

XX
 CC The present invention describes primer sets for synthesizing 5602 full-

XX
 CC length cDNAs defined in the specification. Where a primer set comprises:

XX
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the

XX
 CC complementary strand of a polynucleotide which comprises one of the 5602

XX
 CC nucleotide sequences defined in the specification; or (b) a combination

XX
 CC of an oligonucleotide comprising a sequence complementary to the

XX
 CC complementary strand of a polynucleotide which comprises a 5'-end

XX
 CC sequence and an oligonucleotide comprising a sequence complementary to a

XX
 CC polynucleotide which comprises a 3'-end sequence, where the

XX
 CC oligonucleotide comprises at least 15 nucleotides and the combination of

XX
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the

XX
 CC specification. The primer sets can be used in antisense therapy and in

XX
 CC gene therapy. The primers are useful for synthesizing polynucleotides,

XX
 CC particularly full-length cDNAs. The primers are also useful for the

XX
 CC detection and/or diagnosis of the abnormality of the proteins encoded by

XX
 CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention

XX
 SQ Sequence 349 AA;

Query Match 48.8%; Score 20; DB 4; Length 349;
 Best Local Similarity 100.0%; Pred. No. 1e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NCLYKIFLFDTOHFISKERA 20
 |||||
 DB 202 NCLYKIFLFDTOHFISKERA 221

RESULT 3
 ADE08012
 ID ADE08012 standard; protein; 1149 AA.

XX
 AC ADE08012;

XX
 DT 29-JAN-2004 (first entry)

XX
 DE Novel protein (useful for identifying genetic disorders) #167.

XX
 KM novel gene; novel protein; tissue marker; molecular weight marker;

XX
 KM chromosome marker; genetic disorder.

XX
 OS unidentified.

XX
 PN WO2003054152-A2.

XX
 PD 03-JUL-2003.

XX
 PF 10-DEC-2002; 2002WO-US039555.

XX
 PR 10-DEC-2001; 2001US-039739P.

XX
 PR 11-DEC-2001; 2001US-039453P.

XX
 PR 14-MAR-2002; 2002US-0365091P.

XX
 PR 14-MAR-2002; 2002US-0365384P.

XX
 PR 12-APR-2002; 2002US-0372381P.

XX
 PR 12-APR-2002; 2002US-0372615P.

XX
 PR 22-APR-2002; 2002US-00128558.

XX
 PR 24-APR-2002; 2002US-0376045P.

XX
 PA (HYSE-) HYSEQ INC.

XX
 PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;

XX
 PI Ghosh M, Xie AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;

XX
 PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;

XX
 PI WPI; 2003-569235/53.

XX
 PI N-PDB; ADE07101.

XX
 PS Claim 20; SEQ ID NO 1078; 1177pp; English.

XX
 CC The invention comprises the amino acid and coding sequences of novel

XX
 CC proteins. The DNA and protein sequences of the invention are useful as:

XX
 CC markers for tissues in which the corresponding protein is preferentially

XX
 CC expressed; as molecular weight markers on gels; as chromosome markers or

XX
 CC tags; to identify chromosomes or to map related gene positions; and to

XX
 CC compare with endogenous DNA sequences in patients to identify potential

XX
 CC genetic disorders. The present amino acid sequence represents a protein

XX
 CC of the invention.

XX
 SQ Sequence 1149 AA;

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using BW model

Run on: February 1, 2005, 14:53:01 ; Search time 69.7778 Seconds
(without alignments)
210.782 Million cell updates/sec

Title: SEQ5ALA1854

Sequence: 1 EKMKKMKHMPFMSILQEHIC.....MKKEULTSHQSQTAPFLFA 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2002273 seqs, 358729299 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1316031

Minimum DB seq length: 21
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

1: geneseqp1980s:.*
2: geneseqp1990s:.*
3: geneseqp2000s:.*
4: geneseqp2001s:.*
5: geneseqp2002s:.*
6: geneseqp2003as:.*
7: geneseqp2003bs:.*
8: geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	515	2	AAW54099 Homo sapi
2	20	48.8	349	4	ABD92729 Human pro
3	20	48.8	1149	7	ADB08012 Novel pro
4	20	48.8	2144	4	ABAB5029 Protein e
5	8	19.5	411	7	ABO61395 Klebsiell
6	8	19.5	443	6	ABU42163 Protein e
7	8	19.5	446	6	ABM69250 Phototrab
8	8	19.5	457	7	ADP06196 Bacteri
9	8	19.5	457	6	ABU40177 Protein e
10	8	19.5	473	7	ABO63337 Klebsiell
11	8	19.5	480	7	ADP04247 Bacteri
12	8	19.5	492	7	ABO71657 Pseudom
13	7	17.1	416	7	ADG36882 A. thali
14	7	17.1	417	4	AAAB92908 Human pro
15	7	17.1	431	2	AAV49137 Amino aci
16	7	17.1	588	6	ABU08491 Colled co
17	7	17.1	589	2	AAV43392 Pspc alph
18	7	17.1	605	6	ABU08493 Fragment
19	7	17.1	774	4	AAU14352 Human nov
20	7	17.1	865	6	ABU08489 S. pneumo
21	7	17.1	929	2	AAV44593 Streptoco
22	7	17.1	929	2	AAV43384 S. pneumo
23	7	17.1	1231	6	ABU08490 Fragment
24	6	14.6	41	3	AAAB27686 Human sec
25	6	14.6	45	4	AAO01594 Human pol

26	6	14.6	45	4	AAO04449 Human pol
27	6	14.6	51	3	AAO00262 Human sec
28	6	14.6	56	4	AAU18927 Novel pro
29	6	14.6	56	4	AAU22275 Human car
30	6	14.6	56	4	AAU22275 Human rep
31	6	14.6	56	4	ABG21112 Novel hum
32	6	14.6	56	7	ADK46243 Human car
33	6	14.6	63	4	AAU91234 Human imm
34	6	14.6	74	4	AAO67778 Human foe
35	6	14.6	81	4	AAO08667 Human pol
36	6	14.6	82	4	AAO03775 Human pol
37	6	14.6	83	4	AAO09971 Human pol
38	6	14.6	84	8	ABO56259 Human gen
39	6	14.6	86	5	AAU79243 Human N-m
40	6	14.6	88	4	AAO11546 Human pol
41	6	14.6	88	4	AAO06065 Human pol
42	6	14.6	88	5	ADK36989 Novel hum
43	6	14.6	94	4	AAAB9692 Human imm
44	6	14.6	94	4	AAO02776 Human pol
45	6	14.6	95	4	ABD11340 Human sec

ALIGNMENTS

RESULT 1
AAW54099
ID AAW54099 standard; protein; 515 AA.
XX
AC AAW54099;
XX
DT 28-SEP-1998 (first entry)
XX
DE Homo sapiens BAP28 sequence.
XX
KW BARD1; ring protein; BRCA1; breast cancer; risk; diagnosis.
XX
OS Homo sapiens.
XX
PN MO9812327-A2.
XX
PD 26-MAR-1998.
XX
PF 19-SEP-1997; 97MO-US016842.
XX
PR 20-SEP-1996; 96US-0025296P.
PR 03-APR-1997; 97US-0042611P.
PR 04-APR-1997; 97US-0042985P.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
XX Bowcock AM, Baer R;
XX
XX WPI: 1998-230317/20.
XX N-PSDB; AAV24135.
XX
XX DNA sequence encoding BARD1, B123, BE2, BE14, BE31 or BE445 - which as
XX breast cancer antigen, BRCA1, binding proteins are useful to identify
XX patient having or at risk of developing cancer.
XX
XX Disclosure: Page 287-288; 348pp; English.
XX
XX The sequence is that of a protein which can be used in the preparation of
XX the recombinant breast cancer antigen, BRCA1, binding proteins BARD1,
XX B123, BE2, BE14, BE31 or BE445, or a composition for the detection of a
XX BARD1, B123, BE2, BE14, BE31 or BR445 nucleic acid sequence, specifically
XX a wild type BARD1 composition for the detection or purification of BRCA1,
XX useful to identify a patient having, or at risk of developing cancer.
XX BARD1 can be used in the preparation of an anti-BARD1 antibody, and in
XX the detection and purification of a BRCA1 protein. BARD1, B123, BE2,
XX B14, BE31 or BE445 can be used in the identification of a binding protein
XX agonist or antagonist that alters the binding of BARD1, B123, BE2, BE14,
XX BE31 or BE445 to BRCA1 or the biological activity of the BRCA1-BARD1.

This Page Blank (uspto)

CC B123, BE2, BE14, BE31 or BE445 complex. The antibodies can be used to
CC detect BARD1, B123, BE2, BE14, BE31 or BE445, a specific anti-BARD1
CC antibody can be used to identify a patient having or at risk of
CC developing cancer
XX
SQ Sequence 515 AA:

Query Match 100.0%; Score 41; DB 2; Length 515;
Best Local Similarity 100.0%; Pred. No. 8.6e-35;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKNKNNHGPMSILOEHIGAKKEELTSRHSQSLTAFLTA 41
Db 205 EKNKNNHGPMSILOEHIGAKKEELTSRHSQSLTAFLTA 245

RESULT 2
AAB92729
ID AAB92729 standard; protein, 349 AA.

XX AAB92729;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:11159.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Claim 8; SEQ ID NO 11159; 2537bp + Sequence Listing; English.

CC The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX

SQ Sequence 349 AA;
Query Match 48.8%; Score 20; DB 4; Length 349;
Best Local Similarity 100.0%; Pred. No. 8.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EKNKNNHGPMSILOEHIG 20
Db 39 EKNKNNHGPMSILOEHIG 58

RESULT 3
ADE08012
ID ADE08012 standard; protein, 1149 AA.

XX ADE08012;

XX 29-JAN-2004 (first entry)

XX Novel protein (useful for identifying genetic disorders) #167.

XX novel gene; novel protein; tissue marker; molecular weight marker;

XX chromosome marker; genetic disorder.

XX Unidentified.

XX WO2003054152-A2.

XX 03-JUL-2003.

XX 10-DEC-2002; 2002WO-US039555.

XX 10-DEC-2001; 2001US-0339739P.

XX 11-DEC-2001; 2001US-0339453P.

XX 14-MAR-2002; 2002US-0365091P.

XX 14-MAR-2002; 2002US-0365384P.

XX 12-APR-2002; 2002US-0372381P.

XX 12-APR-2002; 2002US-0372615P.

XX 22-APR-2002; 2002US-00128558.

XX 24-APR-2002; 2002US-0376045P.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QH, Wang J;

XX Ghosh M, Xue AJ, Wehman T, Weng G, Zhou P, Drmanac RT, Wang Z;

XX Ma Y, Wang D, Chen R, Xu C, Boyle BJ;

XX WPI; 2003-569235/53.

XX N-Psdb; ADB07101.

XX New polynucleotides, useful for expressing recombinant proteins for

XX analysis, characterization or therapeutic use, or as markers for tissues

XX in which the corresponding protein is preferentially expressed.

XX Claim 20; SEQ ID NO 1078; 1177bp; English.

XX The invention comprises the amino acid and coding sequences of novel

XX proteins. The DNA and protein sequences of the invention are useful as:

XX markers for tissues in which the corresponding protein is preferentially

XX expressed; as molecular weight markers on gels; as chromosome markers or

XX tags; to identify chromosomes or to map related gene positions; and to

XX compare with endogenous DNA sequences in patients to identify potential

XX genetic disorders. The present amino acid sequence represents a protein

XX of the invention.

XX Sequence 1149 AA;

XX SQ

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

CM protein - protein search, using sw model

Run on: February 1, 2005, 14:53:01 / Search time 69.7776 Seconds
(without alignment)
210.782 Million cell updates/sec

Title: SEQ5ASP1967

Perfect score: 41
Sequence: 1 LKGLFTLPAHLVLPADTL.....QVNISKTDPAFSENDPEK 41

Scoring table: OLIGO

Gapop 60.0, Gapext 60.0

Searched: 2002273 seqs, 35872929 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1316031

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: A_Geneseq_23Sep04.*

1: geneseqp19808.*
2: geneseqp19908.*
3: geneseqp20008.*
4: geneseqp20018.*
5: geneseqp20028.*
6: geneseqp20038.*
7: geneseqp20038.*
8: geneseqp20048.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	515	2	AAW54099 Homo sapi
2	20	48.8	349	4	AA92729 Human pro
3	20	48.8	1149	7	AD808012 Novel pro
4	20	48.8	2144	4	AA85029 Protein e
5	20	48.8	220	2	AA23793 A gyrase
6	17.1	17.1	453	5	ABB55123 Lactococc
7	17.1	17.1	587	7	AB082733 Pseudom
8	17.1	17.1	1045	2	ABP60399 Yeast HMG
9	17.1	17.1	1045	8	AA58611 Yeast HMG
10	17.1	17.1	1045	8	ADM98779 HMG-CoA r
11	17.1	17.1	1045	8	ADM98779 HMG-CoA r
12	17.1	17.1	1045	8	ADM98844 HMG-CoA r
13	17.1	17.1	1045	8	ADM98844 HMG-CoA r
14	17.1	17.1	1045	8	ADM98844 HMG-CoA r
15	17.1	17.1	1045	8	ADM98844 HMG-CoA r
16	17.1	17.1	1045	8	ADM98844 HMG-CoA r
17	17.1	17.1	1045	8	ADM98844 HMG-CoA r
18	17.1	17.1	1045	8	ADM98844 HMG-CoA r
19	17.1	17.1	1045	8	ADM98844 HMG-CoA r
20	17.1	17.1	1045	8	ADM98844 HMG-CoA r
21	17.1	17.1	1045	8	ADM98844 HMG-CoA r
22	17.1	17.1	1045	8	ADM98844 HMG-CoA r
23	17.1	17.1	1045	8	ADM98844 HMG-CoA r
24	17.1	17.1	1045	8	ADM98844 HMG-CoA r
25	17.1	17.1	1045	8	ADM98844 HMG-CoA r

ALIGNMENTS

RESULT 1
AAW54099
ID AAW54099 standard; protein, 515 AA.
AC AAW54099;
DT 28-SEP-1998 (first entry)
XX
DE Homo sapiens BAP28 sequence.
XX
KW BARD1, ring protein; BRCA1, breast cancer; risk; diagnosis.
XX
OS Homo sapiens.
XX
PN WO9812327-A2
XX
PD 26-MAR-1998.
XX
PF 19-SEP-1997; 97WO-US016842.
XX
PR 20-SEP-1996; 96US-005296P.
PR 03-APR-1997; 97US-004261P.
PR 04-APR-1997; 97US-0042985P.
XX
PA (TEXAS) UNIV TEXAS SYSTEM.
PI Bowcock AM, Baer R;
DR WPI: 1998-230317/20.
DR N-PSDB; AAV24135.
XX
PT DNA sequence encoding BARD1, B123, B2, B21, B21 or B2445 - which as
PT breast cancer antigen, BRCA1, binding proteins are useful to identify
PT patient having or at risk of developing cancer.
XX
PS Disclosure; Page 287-288; 348pp; English.
XX
The sequence is that of a protein which can be used in the preparation of
the recombinant breast cancer antigen, BRCA1, binding proteins BARD1,
B123, B2, B21, B21 or B2445, or a composition for the detection of a
BARD1, B123, B2, B21, B21 or B2445 nucleic acid sequence, specifically
a wild type BARD1 composition for the detection or purification of BRCA1,
useful to identify a patient having, or at risk of developing cancer.
CC BARD1 can be used in the preparation of an anti-BARD1 antibody, and in
CC the detection and purification of a BRCA1 protein. BARD1, B123, B2,
CC B21, B21 or B2445 can be used in the identification of a binding protein
CC agonist or antagonist that alters the binding of BARD1, B123, B2, B21,
CC B21 or B2445 to BRCA1 or the biological activity of the BRCA1-BARD1,

This Page Blank (uspto)

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

XX 3

XX 3

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:53:01 ; Search time 69.7778 Seconds
(without alignments)
210.782 Million cell updates/sec

Title: SEQ5GLY2017

Sequence: 1 NCXYKILFPTQHFISKRA.....ALMPLVDQLENRLGSEKLF 41

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 2002273 seqs, 358729239 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1316031

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: A Geneseq_23Sep04:*

1: Geneseqp1980s:*\n2: Geneseqp1990s:*\n3: Geneseqp2000s:*\n4: Geneseqp2001s:*\n5: Geneseqp2002s:*\n6: Geneseqp2003as:*\n7: Geneseqp2003bs:*\n8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	515	2	AAM54099 Homo sapi
2	20	48.8	349	4	AAB92729 Human pro
3	20	48.8	1149	7	ADBE08012 Novel pro
4	20	48.8	2144	4	AAB85029 Protein e
5	6	17.1	118	7	ABM73641 DNA clone
6	7	17.1	650	7	ABO62754 K1ebsle11
7	7	17.1	655	4	AAE13001 Rice poly
8	6	14.6	32	4	ABH15816 Human ner
9	6	14.6	36	4	ABH40457 Peptide #
10	6	14.6	36	4	AAM34177 Peptide #
11	6	14.6	36	4	AAM74002 Human bon
12	6	14.6	36	4	AAM61279 Human bra
13	6	14.6	36	4	ABG55759 Human liv
14	6	14.6	36	5	ABG43897 Human pep
15	6	14.6	36	5	ABH60930 Lung spec
16	6	14.6	52	8	ABO55613 Human gen
17	6	14.6	71	6	ABU40981 Protein e
18	6	14.6	72	5	ABP07304 Human ORF
19	6	14.6	82	4	AAM92148 Human dig
20	6	14.6	83	4	AAM15530 Peptide #
21	6	14.6	83	4	ABB34537 Peptide #
22	6	14.6	83	4	AAM28020 Peptide #
23	6	14.6	83	4	ABB29363 Peptide #
24	6	14.6	83	4	ABB19946 Protein #
25	6	14.6	83	4	AAM67720 Human bon

26	6	14.6	83	4	AAM55323 Human bra
27	6	14.6	83	4	ABG49358 Human liv
28	6	14.6	83	5	ABG37283 Human pep
29	6	14.6	83	5	ADK34692 Novel hum
30	6	14.6	86	4	AAO07992 Human pol
31	6	14.6	87	7	ADM26908 Hyperther
32	6	14.6	95	3	AAO01570 Human sec
33	6	14.6	102	5	AAO17456 Human liv
34	6	14.6	107	4	AAH19971 Human STA
35	6	14.6	108	4	ABG13882 Novel hum
36	6	14.6	138	3	ABH42068 Human ORF
37	6	14.6	138	4	AAO10626 Human pol
38	6	14.6	141	2	AAV07239 Fragment
39	6	14.6	141	2	AAV07240 Fragment
40	6	14.6	141	2	AAV07240 Fragment
41	6	14.6	141	5	ADK36505 Novel hum
42	6	14.6	150	4	ADM19866 Protein e
43	6	14.6	152	4	ABH64818 Drosophi1
44	6	14.6	161	2	AAH77507 Human neu
45	6	14.6	164	4	AAM80897 Human hae

ALIGNMENTS

RESULT 1
AAM54099
ID AAM54099 standard; protein; 515 AA.
XX
AC AAM54099;
DT 28-SEP-1998 (first entry)
XX
DE Homo sapiens BAP28 sequence.
XX
KW BARD1; ring protein; BRCA1; breast cancer; risk; diagnosis.
XX
OS Homo sapiens.
XX
PN NO9812327-A2
PD 26-MAR-1998.
XX
PF 19-SEP-1997; 97WO-US016842.
XX
PR 20-SEP-1996; 96US-0025296P.
PR 03-APR-1997; 97US-0042611P.
PR 04-APR-1997; 97US-0042985P.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Bowcock AM, Baer R;
DR WPI; 1998-230317/20.
XX
DR N-PSDB; AAV24135.
XX
PT DNA sequence encoding BARD1, B123, BE2, BE14, BE31 or BE445 - which as
PT breast cancer antigen, BRCA1, binding proteins are useful to identify
PT patient having or at risk of developing cancer.
XX
PS Disclosure; Page 287-288; 348pp; English.
XX
CC The sequence is that of a protein which can be used in the preparation of
CC the recombinant breast cancer antigen, BRCA1, binding proteins BARD1,
CC B123, BE2, BE14, BE31 or BE445, or a composition for the detection of a
CC BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid sequence, specifically
CC a wild type BARD1 composition for the detection or purification of BRCA1,
CC useful to identify a patient having, or at risk of developing cancer.
CC BARD1 can be used in the preparation of an anti-BARD1 antibody, and in
CC the detection and purification of a BRCA1 protein. BARD1, B123, BE2,
CC BE14, BE31 or BE445 can be used in the identification of a binding protein
CC agonist or antagonist that alters the binding of BARD1, B123, BE2, BE14,
CC BE31 or BE445 to BRCA1 or the biological activity of the BRCA1-BARD1,

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH1633 to AAH18742 represent human cDNA sequences; AA92446 to AA95893
CC represent human amino acid sequences; and AAH1629 to AAH1632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
CX

Query Match 48.8%; Score 20; DB 4; Length 349;
Best Local Similarity 100.0%; Pred. No. 1e-13;

OY
1 NCLYKIFLFDTHFISKERA 20

RESULT 3
ADE08012

12 ADVISORY BOARD; PROBLEM; 1195 MAY
XX

X

[illegible][illegible]

KW chromosome marker; genetic disorder.

OS Unidentified.

PN W02003054152-A2

PD 03-JUL-2003

PF 10-DEC-2002; 2002WO-US039555.

PR 10-DEC-2001; 2001US-0339739P.

PR 14-MAR-2002; 2002US-0365091P.

PR 12-APR-2002; 2002U

PR 22-APR-2002; 2002US-00128558.

XX

[illegible]

PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drymanac RT, Wang Z;

XX	2003	EC000000/EC0
XX	2003	EC000000/EC0

DR N-PSDB; ADE07101.
XX

PT New polynucleotides, useful for expressing recombinant proteins for

PT in which the corresponding protein is preferentially expressed.

PS Claim 20; SEQ ID NO 1078; 1177pp; English.
XX

CC The invention comprises the amino acid and coding sequences of novel
CC proteins. The DNA and protein sequences of the invention are useful in

CC markers for tissues in which the corresponding protein is preferentially
CC expressed: astolocular weight markers on gels as shown above

CC tags; to identify chromosomes or to map related gene positions; and to
CC compare with endogenous DNA sequences in patients to identify potential

CC genetic disorders. The present amino acid sequence represents a protein
CC of the invention

Sequence 1149 AA.

Sequence 1149 AA,

Query Match 48.8%; Score 20; DB 7; Length 1149;
Best Local Similarity 100.0%; Pred. No. 3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 NCLYKIFLPTQHFISKERA 20
Db 1002 NCLYKIFLPTQHFISKERA 1021

RESULT 4
AAB85029 standard; protein; 2144 AA.
XX AAB85029;
AC AAB85029;
XX 06-AUG-2001 (first entry)
XX Protein encoded by BAP28 cDNA consisting of exons 1 to 45.
XX BAP28; prostate; tumour; cancer; diagnostic; genetic analysis.
XX Homo sapiens.
XX Key Location/Qualifiers
FH Misc-difference 1694
FT /label= Ser or Asn
FT Misc-difference 1854
FT /label= Ala or Val
FT Misc-difference 1967
FT /label= Asp or Asn
FT Misc-difference 2017
FT /label= Gly or Glu
XX WO200100669-A2.
XX 04-JAN-2001.
XX 23-JUN-2000; 2000WO-IB001183.
XX 25-JUN-1999; 99US-0141323P.
XX 18-JAN-2000; 2000US-0176880P.
XX (GEST) GENSET.
XX Barry C, Bougueleret L, Chumakov I, Cohen-Akenine A;
XX WPI; 2001-367032/38.
XX N-PSDB; AAF83909, AAF83910.
XX New BAP28 polynucleotides and polypeptides overexpressed in prostate
XX cancer cells for diagnosing prostate tumors, e.g. by hybridization or
XX polymerase chain reaction assays.
XX Claim 14; Page 297-304; 349pp; English.
XX The invention is directed to BAP28 polypeptides, BAP28 polynucleotide
XX sequences and regulatory region located at the 3' and 5' ends of the
XX BAP28 coding region. The BAP28 polypeptides can be expressed by standard
XX recombinant methodology. BAP28 polynucleotides and polypeptides have been
XX found to be over expressed in prostate tumour cells, therefore levels of
XX BAP28 expression and/or activity may be assayed (e.g. by polymerase chain
XX reaction (PCR)) to diagnose patient suffering from or susceptible to
XX prostate cancer. Antibodies specific for the BAP28 polypeptides are
XX useful as diagnostic reagents. Biallelic markers of the BAP28 gene are
XX useful in genetic analysis. The present sequence represents a protein
XX encoded by a first cDNA sequence of the BAP28 gene consisting of the
XX exons 1 to 45
XX
XX Sequence 2144 AA;

Query Match 48.8%; Score 20; DB 4; Length 2144;
Best Local Similarity 100.0%; Pred. No. 5.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NCLYKIFLPTQHFISKERA 20
Db 1997 NCLYKIFLPTQHFISKERA 2016

RESULT 5
ABM73641 standard; protein; 118 AA.
XX ABM73641;
AC ABM73641;
XX 23-OCT-2003 (revised)
XX 17-OCT-2003 (first entry)
XX DNA clone originating in barley containing SNP sequence #51.
XX Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis.
XX Hordeum vulgare; var.
XX WO2003057877-A1.
XX 17-JUL-2003.
XX 16-DEC-2002; 2002WO-IB005403.
XX 20-DEC-2001; 2001JP-00387059.
XX 20-DEC-2001; 2001JP-00387131.
XX 20-DEC-2001; 2001JP-00403299.
XX 20-DEC-2001; 2001JP-00403300.
XX 27-SEP-2002; 2002JP-00327515.
XX (UNINT-) UNIV JAPAN OKAYAMA.
XX Sato K, Takeda K, Kohara Y;
XX WPI; 2003-587127/55.
XX Single nucleotide polymorphism sites in barley varieties and DNA
XX sequences containing them for analysis and identification of barley
XX varieties and production of barley transformants with desired
XX characteristics.
XX Disclosure; SEQ ID XX; 284pp; Japanese.
XX The present invention relates to oligonucleotide clones originating in
XX barley (Hordeum vulgare) which contain single nucleotide polymorphisms
XX (SNP). The oligonucleotides may be used for analysis of SNPs among barley
XX varieties, identification of particular varieties and genotype-phenotype
XX analysis, isolation of specific genes and creation of new varieties by
XX transformation of barley varieties with them and production of new barley
XX varieties with desired properties. The present sequence represents an
XX oligonucleotide clone sequence featured in the specification. The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published-pct-sequences. (Updated on 23-Oct-2003 to
XX standardise OS field)
XX
XX Sequence 118 AA;

Query Match 17.1%; Score 7; DB 7; Length 118;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 26 PLVVDQLE 32
Db 45 PLVVDQLE 51

RESULT 6
ABO62754 standard; protein; 650 AA.
ID ABO62754

```

XX ABO62754;
XX
XX 29-JUL-2004 (first entry)
XX
XX Klebsiella pneumoniae polypeptide seqid 9271.
XX
XX Recombinant expression vector; transcription regulatory element;
XX Klebsiella pneumoniae protein; antibacterial; vaccine.
XX
XX Klebsiella pneumoniae.
XX
XX US610836-B1.
XX
XX 26-AUG-2003.
XX
XX 27-JAN-2000; 2000US-00489039.
XX
XX 29-JAN-1999; 99US-0117747P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Breton GL, Osborne M;
XX
XX WPI; 2003-895346/82.
XX
XX N-PSDB; ACH96305.
XX
XX New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
XX preparing a vaccine composition against Klebsiella pneumoniae.
XX
XX Disclosure; SEQ ID NO 9271; 932pp; English.
XX
XX
XX The invention describes a new isolated nucleic acid encoding a Klebsiella
XX pneumoniae polypeptide. Also described are: a recombinant expression
XX vector comprising the nucleic acid, operably linked to a transcription
XX regulatory element; and a cell comprising the recombinant expression
XX vector. The nucleic acid is useful for preparing a vaccine composition
XX against Klebsiella pneumoniae. This is the amino acid sequence of a
XX Klebsiella pneumoniae polypeptide of the invention
XX
XX Sequence 650 AA;
XX
XX
XX Query Match 17.1%; Score 7; DB 7; Length 650;
XX Best Local Similarity 100.0%; Pred. No. 55;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 21 GATMMP 27
XX 9 GATMMP 15
XX
XX Db
XX
XX RESULT 7
XX AAE13001
XX ID AAE13001 standard; protein; 655 AA.
XX
XX AC AAE13001;
XX
XX 28-JAN-2002 (first entry)
XX
XX Rice poly (A) binding protein from clone rrr9n.pk005.117.
XX
XX Rice; polyadenylated RNA-binding protein; transgenic plant; herbicide;
XX eukaryotic translation initiation factor-4 (eif-4) gamma;
XX genetic mapping; physical mapping.
XX
XX Oryza sativa.
XX
XX US6294658-B1.
XX
XX 25-SEP-2001.
XX
XX 02-JUL-1999; 99US-00347833.
XX
XX

```

```

PR 10-JUL-1998; 98US-0092415P.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX Famodu LO, Odell JT;
XX
XX WPI; 2001-656175/75.
XX
XX N-PSDB; AAD21314.
XX
XX New polynucleotides encoding poly(A) binding protein or a eukaryotic
XX translation initiation factor-4 gamma, useful for creating transgenic
XX plants where the proteins are present at higher or lower levels than
XX normal.
XX
XX Claim 9; Fig 1-3; 26pp; English.
XX
XX
XX The present invention relates to an isolated polynucleotide encoding a
XX polyadenylated RNA-binding protein or eukaryotic translation initiation
XX factor-4 (eif-4) gamma. The nucleic acid fragments may be used in PCR
XX protocols to amplify longer nucleic acid fragments encoding homologous
XX genes from DNA or RNA, to create transgenic plants in which the new
XX polypeptides are present at higher or lower levels than normal or in cell
XX types or developmental stages in which they are not normally found. The
XX nucleic acids may also be used as probes for genetic or physical mapping
XX the genes that they are part of and as markers for traits linked to those
XX genes. The polypeptides may be used to produce antibodies and as targets
XX that can be used as herbicides. The identification of inhibitors of those enzymes
XX binding protein from clone rrr9n.pk005.117
XX
XX Sequence 655 AA;
XX
XX
XX Query Match 17.1%; Score 7; DB 4; Length 655;
XX Best Local Similarity 100.0%; Pred. No. 55;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 26 PLVDOLE 32
XX 583 PLVDOLE 589
XX
XX Db
XX
XX RESULT 8
XX ABB15816
XX ID ABB15816 standard; protein; 32 AA.
XX
XX AC ABB15816;
XX
XX 23-JAN-2002 (first entry)
XX
XX Human nervous system related polypeptide SEQ ID NO 4473.
XX
XX Human; nootropic; neuroprotective; cytosolic; dermatological; virucide;
XX immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnerary;
XX antiparkinsonian; antistroke; antianemic; antidiabetic; cancer;
XX antithrombotic; hepatoprotective; cerebroprotective; antiinflammatory;
XX antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
XX antiparasitic; cardiac; immune disorder; cardiovascular disorder;
XX neurological disease; infection; nephrotropic; gene therapy; vaccine.
XX
XX Homo sapiens.
XX
XX WO200159063-A2.
XX
XX 16-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001334.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX
XX 04-FEB-2000; 2000US-0180628P.
XX
XX 24-FEB-2000; 2000US-0184664P.
XX
XX 02-MAR-2000; 2000US-0186350P.
XX
XX 16-MAR-2000; 2000US-0189874P.
XX
XX 17-MAR-2000; 2000US-0190076P.
XX

```


CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
 CC ovarian cancer and other cancers of the adrenal gland, bone marrow,
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune hemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC cardiovascular disorders such as myocardial ischaemia, (d) wound healing
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 CC
 XX Sequence 32 AA;
 SQ
 Query Match 14.6%; Score 6; DB 4; Length 32;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 4 YKIFLP 9
 Db 21 YKIFLP 26
 RESULT 9
 ABB40457
 ID ABB40457 standard; peptide; 36 AA.
 XX ABB40457;
 AC
 XX
 DT 04-FEB-2002 (first entry)
 XX
 DE Peptide #7963 encoded by human foetal liver single exon probe.
 XX
 KM Human; foetal liver; gene expression; single exon nucleic acid probe.
 XX
 OS Homo sapiens.
 XX
 PN WO200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000669.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 XX
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for analyzing
 XX gene expression in human foetal liver.
 XX
 PS Claim 27; SEQ ID NO 33092; 639bp + Sequence Listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for measuring
 CC human gene expression in a sample derived from human foetal liver. The
 CC single exon nucleic acid probes may be used for predicting, measuring and
 CC displaying gene expression in samples derived from human foetal liver. The
 CC present sequence is a peptide encoded by a single exon nucleic acid probe
 CC of the invention. Note: The sequence data for this patent did not form
 CC part of the printed specification, but was obtained in electronic format

CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 36 AA;
 Query Match 14.6%; Score 6; DB 4; Length 36;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 36 GGEERF 41
 Db 4 GGEERF 9
 RESULT 10
 AAM34177
 ID AAM34177 standard; protein; 36 AA.
 XX AAM34177;
 AC
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Peptide #8214 encoded by probe for measuring placental gene expression.
 XX
 KM Probe; microarray; human; placenta; antenatal diagnosis;
 XX genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000663.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488897/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for analyzing
 XX gene expression in human placenta.
 XX
 PS Claim 27; SEQ ID NO 34446; 654bp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENP;
 CC see A131315-A157546). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders
 XX
 SQ Sequence 36 AA;
 Query Match 14.6%; Score 6; DB 4; Length 36;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 36 GGEERF 41
 Db 4 GGEERF 9
 RESULT 11
 AAM74002

ID AAM74002 standard; protein; 36 AA.
XX
AC AAM74002;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 34308.
XX
KM Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000668.
XX
PR 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX
PS Example 4; SEQ ID NO 34308; 658bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention
XX
SQ Sequence 36 AA;
XX
Query Match 14.6%; Score 6; DB 4; Length 36;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
DY 36 GGEKEF 41
XX |||||
DB 4 GGEKEF 9
XX
RESULT 12
AAM61279
ID AAM61279 standard; protein; 36 AA.
XX
AC AAM61279;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 33384.
XX
KM Human; brain expressed exon; gene expression analysis; probe; microarray;
XX Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX

PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000667.
XX
PR 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX
PS Example 4; SEQ ID NO 33384; 650bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention
XX
SQ Sequence 36 AA;
XX
Query Match 14.6%; Score 6; DB 4; Length 36;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
DY 36 GGEKEF 41
XX |||||
DB 4 GGEKEF 9
XX
RESULT 13
ABG55759
ID ABG55759 standard; peptide; 36 AA.
XX
AC ABG55759;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver peptide; SEQ ID NO 34407.
XX
KM Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX hypercholesterolaemia; coronary heart disease.
XX
OS Homo sapiens.
XX
PN WO200157273-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000664.
XX
PR 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX

XX
FM (MOLE-) MOLECULAR DYNAMICS INC.
XX

KW non-cancerous diseases of the lung: transgenic animal

OS Homo sapiens.
 XX
 PN WO200268633-A2.
 XX
 PD 06-SEP-2002.
 XX
 PF 21-NOV-2001; 2001WO-US043612.
 XX
 PR 22-NOV-2000; 2000US-0252500P.
 XX
 PA (DIAD-) DIADEXUS INC.
 XX
 PI Macina RA, Recipon H, Chen S, Sun Y, Liu C;
 XX
 DR WPI; 2002-713376/77.
 XX
 PT New isolated human nucleic acid molecule and polypeptide, useful for
 PT identifying, diagnosing, monitoring, staging, imaging and treating lung
 XX cancer and non-cancerous diseases of the lung.
 XX
 PS Claim 11; Page 328; 389pp; English.
 CC
 CC The invention describes an isolated human nucleic acid (1) encoding any
 CC of 120 10-1533 residue amino acid sequences (S1), given in the
 CC specification, comprising any of 164 179-12421 base pair sequences (S2),
 CC given in the specification. The methods and compositions of the present
 CC invention are useful for identifying, diagnosing, monitoring, staging,
 CC imaging and treating lung cancer and non-cancerous diseases of the lung.
 CC They are also used for identifying lung tissue, monitoring and
 CC identifying and/or designing antagonists of the polypeptide of the
 CC invention, gene therapy, production of transgenic animals and production
 CC of engineered lung tissue for treatment and research. This is the amino
 CC acid sequence of a lung specific nucleic acid
 XX
 SQ Sequence 48 AA;

Query Match 14.6%; Score 6; DB 5; Length 48;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 PLVDQL 31
 |||||
 Db 42 PLVDQL 47

Search completed: February 1, 2005, 15:20:30
 Job time : 82.7778 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OW protein - protein search, using sw model

Run on: February 1, 2005, 14:53:01 / Search time 69.7778 Seconds
(without alignments)
210.782 Million cell updates/sec

Title: SEQ5ALA1854

Sequence: 1 EKMKWKNHMFMSILQEHIG.....MKKELTSHQSQTAFLEA 41

Scoring table: OLIGO

Gapop 60.0, Gapext 60.0

Searched: 2002273 seqs, 358729239 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1316031

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: A_Geneseq_23Sep04:*

1: geneseqp1980s:*\n2: geneseqp1990s:*\n3: geneseqp2000s:*\n4: geneseqp2001s:*\n5: geneseqp2002s:*\n6: geneseqp2003as:*\n7: geneseqp2003bs:*\n8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	515	2	AAW54099 Homo sapi
2	20	48.8	349	4	AAW54099 Homo sapi
3	20	48.8	1149	7	AAW54099 Homo sapi
4	20	48.8	2144	4	AAW54099 Homo sapi
5	8	19.5	411	7	AAW54099 Homo sapi
6	8	19.5	443	6	AAW54099 Homo sapi
7	8	19.5	446	6	AAW54099 Homo sapi
8	8	19.5	457	7	AAW54099 Homo sapi
9	8	19.5	467	6	AAW54099 Homo sapi
10	8	19.5	473	7	AAW54099 Homo sapi
11	8	19.5	480	7	AAW54099 Homo sapi
12	8	19.5	492	7	AAW54099 Homo sapi
13	8	19.5	497	7	AAW54099 Homo sapi
14	8	19.5	507	7	AAW54099 Homo sapi
15	8	19.5	515	2	AAW54099 Homo sapi
16	8	19.5	515	2	AAW54099 Homo sapi
17	8	19.5	515	2	AAW54099 Homo sapi
18	8	19.5	515	2	AAW54099 Homo sapi
19	8	19.5	515	2	AAW54099 Homo sapi
20	8	19.5	515	2	AAW54099 Homo sapi
21	8	19.5	515	2	AAW54099 Homo sapi
22	8	19.5	515	2	AAW54099 Homo sapi
23	8	19.5	515	2	AAW54099 Homo sapi
24	8	19.5	515	2	AAW54099 Homo sapi
25	8	19.5	515	2	AAW54099 Homo sapi

26	6	14.6	45	4	AAW54099 Homo sapi
27	6	14.6	51	3	AAW54099 Homo sapi
28	6	14.6	56	4	AAW54099 Homo sapi
29	6	14.6	56	4	AAW54099 Homo sapi
30	6	14.6	56	4	AAW54099 Homo sapi
31	6	14.6	56	4	AAW54099 Homo sapi
32	6	14.6	56	4	AAW54099 Homo sapi
33	6	14.6	56	4	AAW54099 Homo sapi
34	6	14.6	56	4	AAW54099 Homo sapi
35	6	14.6	56	4	AAW54099 Homo sapi
36	6	14.6	56	4	AAW54099 Homo sapi
37	6	14.6	56	4	AAW54099 Homo sapi
38	6	14.6	56	4	AAW54099 Homo sapi
39	6	14.6	56	4	AAW54099 Homo sapi
40	6	14.6	56	4	AAW54099 Homo sapi
41	6	14.6	56	4	AAW54099 Homo sapi
42	6	14.6	56	4	AAW54099 Homo sapi
43	6	14.6	56	4	AAW54099 Homo sapi
44	6	14.6	56	4	AAW54099 Homo sapi
45	6	14.6	56	4	AAW54099 Homo sapi

ALIGNMENTS

RESULT 1	AAW54099	AAW54099 standard; protein: 515 AA.
ID	AAW54099	
AC	AAW54099	
DT	28-SEP-1998	(first entry)
XX		
DB	Homo sapiens BAP28 sequence.	
XX		
KW	BARD1, ring protein, BRCA1, breast cancer, risk, diagnosis.	
XX		
OS	Homo sapiens.	
XX		
PN	W09812327-A2	
XX		
PD	26-MAR-1998.	
XX		
PF	19-SEP-1997; 97WO-US016842.	
XX		
PR	20-SEP-1996; 96US-0025296P.	
PR	03-APR-1997; 97US-0042611P.	
PR	04-APR-1997; 97US-0042985P.	
XX		
PA	(TEXA) UNIV TEXAS SYSTEM.	
XX		
FI	Bowcock AM, Baer R;	
XX		
DR	WPI; 1998-230317/20.	
DR	N-PSDB; AAV24135.	
XX		
PT	DNA sequence encoding BARD1, B123, BE2, BE14, BE31 or BE445 - which as	
PT	patient cancer antigen, BRCA1, binding proteins are useful to identify	
XX		
PS	disclosure; Page 287-288; 348pp; English.	
XX		
CC	The sequence is that of a protein which can be used in the preparation of	
CC	the recombinant breast cancer antigen, BRCA1, binding proteins BARD1,	
CC	B123, BE2, BE14, BE31 or BE445, or a composition for the detection of a	
CC	BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid sequence, specifically	
CC	a wild type BARD1 composition for the detection or purification of BRCA1,	
CC	useful to identify a patient having, or at risk of developing cancer.	
CC	BARD1 can be used in the preparation of an anti-BARD1 antibody, and in	
CC	the detection and purification of a BRCA1 protein. BARD1, B123, BE2,	
CC	BE14, BE31 or BE445 can be used in the identification of a binding protein	
CC	antigen or antagonist that alters the binding of BARD1, B123, BE2, BE14,	
CC	BE31 or BE445 to BRCA1 or the biological activity of the BRCA1-BARD1,	

CC B123, BE2, BE14, BE31 or BE445 complex. The antibodies can be used to
 CC detect BAP1, B123, BE2, BE14, BE31 or BE445, a specific anti-BAP1
 CC antibody can be used to identify a patient having or at risk of
 CC developing cancer
 CC
 SQ Sequence 515 AA;

Query Match 100.0%; Score 41; DB 2; Length 515;
 Best Local Similarity 100.0%; Pred. No. 8,6e-35;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKNMKNHMGPFMSITQEHIGAMKKEELTSSQSLTAPFLA 41
 DB 205 EKNMKNHMGPFMSITQEHIGAMKKEELTSSQSLTAPFLA 245

RESULT 2
 AAB92729
 ID AAB92729 standard; protein; 349 AA.
 AC AAB92729;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:11159.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 FN EPI074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-00116126.
 XX
 PR 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.
 XX
 PS Claim 8; SEQ ID NO 11159; 2537bp + Sequence Listing; English.

CC The present invention describes primer sets for synthesizing 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX
 SQ Sequence 349 AA;

Query Match 48.8%; Score 20; DB 4; Length 349;
 Best Local Similarity 100.0%; Pred. No. 8.5e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKNMKNHMGPFMSITQEHIG 20
 DB 39 EKNMKNHMGPFMSITQEHIG 58

RESULT 3
 ADE08012
 ID ADE08012 standard; protein; 1149 AA.
 AC ADE08012;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Novel protein (useful for identifying genetic disorders) #167.
 XX
 KW novel gene; novel protein; tissue marker; molecular weight marker;
 KW chromosome marker; genetic disorder.
 XX
 OS Unidentified.
 XX
 FN WO2003054152-A2.
 XX
 PD 03-JUL-2003.
 XX
 PF 10-DEC-2002; 2002WO-US039555.
 XX
 PR 10-DEC-2001; 2001US-0397339P.
 PR 11-DEC-2001; 2001US-0339453P.
 PR 14-MAR-2002; 2002US-0365091P.
 PR 14-MAR-2002; 2002US-0365384P.
 PR 12-APR-2002; 2002US-0372381P.
 PR 12-APR-2002; 2002US-0372615P.
 PR 22-APR-2002; 2002US-00128558.
 PR 24-APR-2002; 2002US-0376045P.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
 PI Ghosh M, Xue AJ, Wehman T, Weng G, Zhou P, Drmanac RT, Wang Z;
 PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ,
 XX
 DR WPI; 2003-569235/53.
 DR N-PSDB; ADE07101.

PT New polynucleotides, useful for expressing recombinant proteins for
 PT analysis, characterization or therapeutic use, or as markers for tissues
 PT in which the corresponding protein is preferentially expressed.
 XX
 PS Claim 20; SEQ ID NO 1078; 1177bp; English.

CC The invention comprises the amino acid and coding sequences of novel
 CC proteins. The DNA and protein sequences of the invention are useful as:
 CC markers for tissues in which the corresponding protein is preferentially
 CC expressed; as molecular weight markers on gels; as chromosome markers or
 CC tags; to identify chromosomes or to map related gene positions; and to
 CC compare with endogenous DNA sequences in patients to identify potential
 CC genetic disorders. The present amino acid sequence represents a protein
 CC of the invention.
 XX
 SQ Sequence 1149 AA;

Query Match 48.8%; Score 20; DB 7; Length 1149;
 Best Local Similarity 100.0%; Pred. No. 2,2e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 MKKEELTSHQSOLTAFFLEA 41
 |||||
 DB 860 MKKEELTSHQSOLTAFFLEA 879

RESULT 4

AAB85029 standard; protein; 2144 AA.

XX AAB85029;

XX 06-AUG-2001 (first entry)

XX Protein encoded by BAP28 cDNA consisting of exons 1 to 45.

XX BAP28; prostate; tumour; cancer; diagnostic; genetic analysis.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Misc-difference 1694 /label= Ser or Asn

XX Misc-difference 1854 /label= Ala or Val

XX Misc-difference 1967 /label= Asp or Asn

XX Misc-difference 2017 /label= Gly or Glu

XX WO200100669-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-1B001183.

XX 25-JUN-1999; 99US-0141323P.

XX 18-JAN-2000; 2000US-017680P.

XX (GIST) GENSET.

XX Barry C, Bougueleret L, Chumakov I, Cohen-Akenine A;

XX WPI; 2001-367032/38.

XX N-PSDB; AAF83909, AAF83910.

XX New BAP28 polynucleotides and polypeptides overexpressed in prostate

XX cancer cells for diagnosing prostate tumors, e.g. by hybridization or

XX polymerase chain reaction assays.

XX Claim 14; Page 297-304; 349pp; English.

XX The invention is directed to BAP28 polypeptides, BAP28 polynucleotide
 CC sequences and regulatory region located at the 3' and 5' ends of the
 CC BAP28 coding region. The BAP28 polypeptides can be expressed by standard
 CC recombinant methodology. BAP28 polynucleotides and polypeptides have been
 CC found to be over expressed in prostate tumour cells, therefore levels of
 CC BAP28 expression and/or activity may be assayed (e.g. by polymerase chain
 CC reaction (PCR)) to diagnose patient suffering from or susceptible to
 CC prostate cancer. Antibodies specific for the BAP28 polypeptides are
 CC useful as diagnostic reagents. Biallelic markers of the BAP28 gene are
 CC useful in genetic analysis. The present sequence represents a protein
 CC encoded by a first cDNA sequence of the BAP28 gene consisting of the
 CC exons 1 to 45

XX Sequence 2144 AA;

Query Match 48.8%; Score 20; DB 4; Length 2144;
 Best Local Similarity 100.0%; Pred. No. 3.6e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKXKNMGPPMSILOEHIG 20
 |||||
 DB 1834 EKXKNMGPPMSILOEHIG 1853

RESULT 5

ABO61395 standard; protein; 411 AA.

XX ABO61395;

XX 29-JUL-2004 (first entry)

XX Klebsiella pneumoniae polypeptide seqid 7912.

XX Recombinant expression vector; transcription regulatory element;

XX Klebsiella pneumoniae protein; antibacterial; vaccine.

XX Klebsiella pneumoniae.

XX US6610836-B1.

XX 26-AUG-2003.

XX 27-JAN-2000; 2000US-00489039.

XX 29-JAN-1999; 99US-0117747P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Breton GL, Osborne M;

XX WPI; 2003-895346/82.

XX N-PSDB; ACH94946.

XX New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for

XX preparing a vaccine composition against Klebsiella pneumoniae.

XX Disclosure; SEQ ID NO 7912; 932pp; English.

XX The invention describes a new isolated nucleic acid encoding a Klebsiella

XX pneumoniae polypeptide. Also described are: a recombinant expression

XX vector comprising the nucleic acid, operably linked to a transcription

XX regulatory element; and a cell comprising the recombinant expression

XX vector. The nucleic acid is useful for preparing a vaccine composition

XX against Klebsiella pneumoniae. This is the amino acid sequence of a

XX Klebsiella pneumoniae polypeptide of the invention

XX Sequence 411 AA;

XX Query Match 19.5%; Score 8; DB 7; Length 411;

XX Best Local Similarity 100.0%; Pred. No. 4.3;

XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX OY 34 LTAFFLEA 41
 |||||

XX DB 46 LTAFFLEA 53

XX RESULT 6

XX ID ABO42163 standard; protein; 443 AA.

XX AC ABO42163;

XX DT 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #27690.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX KW Pseudomonas syringae.

XX OS

XX WO20027183-A2.
 PN 03-OCT-2002.
 PD
 XX 21-MAR-2002; 2002WO-US009107.
 PF
 XX 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW,
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH,
 XX WPI: 2003-029926/02.
 DR N-PSDB; ACMA6033.
 DR
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 70087; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 443 AA;
 Query Match 19.5%; Score 8; DB 6; Length 443;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 34 LTAFFLEA 41
 |||||
 DB 67 LTAFFLEA 74

XX ABM69250;
 AC
 XX 20-NOV-2003 (first entry)
 DT
 XX Photobhabdus luminescens protein sequence #2347.
 DE
 XX
 XX Antibacterial, fungicide, insecticide, polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 KW whooping cough.
 XX
 OS Photobhabdus luminescens.
 XX
 XX WO200294867-A2.
 PN
 XX 28-NOV-2002.
 PD
 XX 07-FEB-2002; 2002WO-IB003040.
 PF
 XX 07-FEB-2001; 2001FR-00001659.
 PR
 XX (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunet F, Danchin A,
 PI Buchrieser C,
 XX WPI: 2003-148459/14.
 DR
 XX Genomic sequence of Photobhabdus luminescens and encoded polypeptides,
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
 PT
 XX Claim 2; SEQ ID NO 2347; 1205pp; French.
 PS
 XX
 CC The invention relates to the isolation of genes and their encoded
 CC proteins from Photobhabdus luminescens. The isolated sequences are
 CC sources of probes and primers for detecting the genome of *P. luminescens*
 CC and related species; to study polymorphisms; for gene analysis and for
 CC detection/amplification of the genes. Antibodies (Ab) raised against the
 CC of *P. luminescens*, e.g. in foods. The genes, proteins, Ab and cells that
 CC carry a gene-containing vector are used to select compounds that
 CC modulate, regulate, induce or inhibit expression of the genes in plants,
 CC animals or microorganisms other than *P. luminescens* and are able to alter
 CC response or sensitivity to toxins and antibiotics produced by *P.*
 CC luminescens. Cells transformed to express the genes are useful for
 CC recombinant production of the proteins, particularly toxins and
 CC antibacterials useful as insecticides, bactericides and fungicides. The
 CC genes, proteins, vectors containing the genes and Ab are also useful
 CC therapeutically (to treat microbial infection by bacteria or fungi that
 CC are sensitive to *P. luminescens*-encoded toxins or antibiotics) and as
 CC biopesticides. Other uses of the genes and the proteins are as virulence
 CC factors and for identifying targets of human diseases for which *P.*
 CC luminescens is a model (particularly plague and whooping cough). This
 CC sequence represents one of the isolated *P. luminescens* proteins.
 XX
 SQ Sequence 446 AA;
 Query Match 19.5%; Score 8; DB 6; Length 446;
 Best Local Similarity 100.0%; Pred. No. 4.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 34 LTAFFLEA 41
 |||||
 DB 103 LTAFFLEA 110

RESULT 8
 ADF06196
 ID ADF06196 standard; protein; 457 AA.
 XX
 AC ADF06196;

```

XX 12-FEB-2004 (first entry)
DT
XX Bacterial polypeptide #2309.
DE
XX Proteus mirabilis infection; bacterial infection; antibacterial;
KM immunostimulant.
XX
XX Proteus mirabilis.
OS
XX US6605709-B1.
PN
XX 12-AUG-2003.
PD
XX 05-APR-2000; 2000US-00543681.
PF
XX 09-APR-1999; 99US-0128706P.
PR
XX (GENO-) GENOME THERAPEUTICS CORP.
PA
XX Breton GL;
PI
XX WPI; 2003-895291/82.
DR
XX N-PSDB; ADF02024.
DR
XX
XX New Proteus mirabilis polypeptides and polynucleotides, useful as
PT reagents for diagnosis of bacterial disease, as components of
PT antibacterial vaccines, as targets for antibacterial drugs, or as
PT biocontrol agents for plants.
XX
XX Disclosure; SEQ ID NO 6481; 870pp; English.
PS
XX
CC The invention relates to new Proteus mirabilis polypeptides and
CC polynucleotides. The invention also relates to antibodies against the
CC polypeptides, methods for producing the polypeptides, a method of
CC generating vaccines for immunising an individual against P. mirabilis, a
CC method for evaluating a compound for the ability to bind a P. mirabilis
CC polypeptide and a method for screening test compounds for anti-bacterial
CC activity. The polypeptides and polynucleotides are useful as molecular
CC targets for diagnosing, preventing and treating pathological conditions
CC resulting from bacterial infection, as reagents for diagnosis of
CC bacterial diseases, as components of antibacterial vaccines, as targets
CC for antibacterial drugs or as bio-control agents for plants. This
CC sequence represents a Proteus mirabilis polypeptide of the invention.
XX
SQ Sequence 457 AA;
Query Match 19.5%; Score 8; DB 7; Length 457;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 34 LTAFFLEA 41
DB 110 LTAFFLEA 117

```

```

XX 21-MAR-2002; 2002WO-US009107.
PF
XX
XX 21-MAR-2001; 2001US-00915242.
PR
XX 06-SEP-2001; 2001US-00948893.
PR
XX 25-OCT-2001; 2001US-0342923P.
PR
XX 08-FEB-2002; 2002US-00072851.
PR
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
DR
XX N-PSDB; ACA44047.
DR
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 68101; 1766pp; English.
PS
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 467 AA;
Query Match 19.5%; Score 8; DB 6; Length 467;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 34 LTAFFLEA 41
DB 92 LTAFFLEA 99

```

```

RESULT 10
ABO63337
ID ABO63337 standard; protein; 473 AA.
AC ABO63337;
XX
XX 29-JUN-2004 (first entry)
DT

```

```

XX Klebsiella pneumoniae polypeptide seqid 9854.
XX
XX Recombinant expression vector; transcription regulatory element;
KM Klebsiella pneumoniae protein; antibacterial; Vaccine.
XX
XX Klebsiella pneumoniae.
XX
XX US6610836-B1.
XX
XX 26-AUG-2003.
XX
XX 27-JUN-2000; 2000US-00489039.
XX
XX 29-JAN-1999; 99US-0117747P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Breton GL, Osborne M,
XX
XX WPI; 2003-895346/82.
XX
XX N-PSDB; ACH96888.
XX
XX New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
PT preparing a vaccine composition against Klebsiella pneumoniae.
XX
XX Disclosure; SEQ ID NO 9854; 932pp; English.
XX
XX The invention describes a new isolated nucleic acid encoding a Klebsiella
CC pneumoniae polypeptide. Also described are: a recombinant expression
CC vector comprising the nucleic acid, operably linked to a transcription
CC regulatory element; and a cell comprising the recombinant expression
CC vector. The nucleic acid is useful for preparing a vaccine composition
CC against Klebsiella pneumoniae. This is the amino acid sequence of a
CC Klebsiella pneumoniae polypeptide of the invention
SQ
Sequence 473 AA;

Query Match      19.5%; Score 8; DB 7; Length 473;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      34 LTAFFLEA 41
      |||||
      111 LTAFFLEA 118

Db

RESULT 11
ADF04247
ID ADF04247 standard; protein; 480 AA.
XX
XX ADF04247;
XX
XX 12-FEB-2004 (first entry)
XX
XX Bacterial polypeptide #360.
DE
XX
XX Proteus mirabilis infection; bacterial infection; antibacterial;
KW immunostimulant.
XX
XX Proteus mirabilis.
OS
XX US6605709-B1.
XX
XX 12-AUG-2003.
XX
XX 05-APR-2000; 2000US-00543681.
XX
XX 09-APR-1999; 99US-0128706P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Breton GL;

```

```

XX WPI; 2003-895291/82.
XX
XX N-PSDB; ADF00075.
XX
XX New Proteus mirabilis polypeptides and polynucleotides, useful as
PT reagents for diagnosis of bacterial disease, as components of
PT antibacterial vaccines, as targets for antibacterial drugs, or as
XX biocontrol agents for plants.
XX
XX Disclosure; SEQ ID NO 4532; 870pp; English.
XX
XX The invention relates to new Proteus mirabilis polypeptides and
CC polynucleotides. The invention also relates to antibodies against the
CC polypeptides, methods for producing the polypeptides, a method of
CC generating vaccines for immunising an individual against P. mirabilis, a
CC method for evaluating a compound for the ability to bind a P. mirabilis
CC polypeptide and a method for screening test compounds for antibacterial
CC activity. The polypeptides and polynucleotides are useful as molecular
CC targets for diagnosing, preventing and treating pathological conditions
CC resulting from bacterial infection, as reagents for diagnosis of
CC bacterial diseases, as components of antibacterial vaccines, as targets
CC for antibacterial drugs or as bio-control agents for plants. This
CC sequence represents a Proteus mirabilis polypeptide of the invention.
XX
XX Sequence 480 AA;

Query Match      19.5%; Score 8; DB 7; Length 480;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      34 LTAFFLEA 41
      |||||
      139 LTAFFLEA 146

Db

RESULT 12
ABO71657
ID ABO71657 standard; protein; 492 AA.
XX
XX ABO71657;
XX
XX 29-JUL-2004 (first entry)
XX
XX Pseudomonas aeruginosa polypeptide #3832.
DE
XX
XX Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.
KW
XX
XX Pseudomonas aeruginosa.
OS
XX US6551795-B1.
XX
XX 22-APR-2003.
XX
XX 18-FEB-1999; 99US-00252991.
XX
XX 18-FEB-1998; 98US-0074788P.
XX
XX 27-JUL-1998; 98US-0094190P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX
XX WPI; 2003-615309/58.
XX
XX N-PSDB; ABD05228.
XX
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
XX Disclosure; SEQ ID NO 20403; 455pp; English.
XX
XX The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and

```

CC therapy of pathological conditions, as molecular targets for diagnostics,
 CC prophylaxis and treatment of pathological conditions resulting from a
 CC bacterial infection, for evaluating a compound, such as a polypeptide,
 CC for the ability to bind a P. aeruginosa nucleic acid, as components of
 CC effective antibacterial targets, as targets for antibacterial drugs,
 CC including anti-P. aeruginosa drugs, as templates for recombinant
 CC production of P. aeruginosa-derived peptides or polypeptides, as target
 CC components for diagnosis and/or treatment of P. aeruginosa-caused
 CC infection, and in detection of P. aeruginosa sequences or other sequences
 CC of Pseudomonas species using biochip technology. Sequences ABO67826-
 CC AB084396 represent P. aeruginosa polypeptides of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format from USPTO at
 CC seqdata.uspto.gov/sequence.html
 XX
 SQ Sequence 492 AA;

Query Match 19.5%; Score 8; DB 7; Length 492;
 Best Local Similarity 100.0%; Pred. No. 4.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 |||||
 DB 107 LTAFFLEA 114

RESULT 13
 ADG36882
 ID ADG36882 standard; protein; 416 AA.

XX ADG36882;

DT 26-FEB-2004 (first entry)

DE A. thaliana MID1B protein.

KM MID1B; extension-activated calcium permeable channel; plant; gravity;

KM tropism; gravity sensor; photosynthesis.

XX Arabidopsis thaliana.

PN JP2003180367-A.

PD 02-JUL-2003.

PF 19-DEC-2001; 2001JP-00385513.

PR 19-DEC-2001; 2001JP-00385513.

PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.

PA (RIKA) RIKAKU KENKYUSHO.

DR WPI; 2003-819302/77.

DR N-PSDB; ADG36881.

PT Novel AtMID1B gene derived from Arabidopsis thaliana encoding extension-

PT tropism in plant.

XX Claim 2; SEQ ID NO 2; 11pp; Japanese.

CC This invention describes a novel gene (AtMID1B) encoding an extension-
 CC activated calcium (Ca²⁺) permeable channel of higher plant, useful for
 CC controlling gravity tropism in plants. The gene is useful as gravity
 CC sensor in plant. Transgenic plants containing the gene have improved
 CC photosynthetic ability.
 CC
 XX

SQ Sequence 416 AA;

Query Match 17.1%; Score 7; DB 7; Length 416;
 Best Local Similarity 100.0%; Pred. No. 4.9;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 KXELTS 29
 DB 243 KXELTS 249

RESULT 14
 AAB92908
 ID AAB92908 standard; protein; 417 AA.

XX AAB92908;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:11537.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EPI074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-00116126.

PR 29-JUL-1999; 99JP-00248036.

PR 27-AUG-1999; 99JP-00300253.

PR 11-JAN-2000; 2000JP-00118776.

PR 02-MAY-2000; 2000JP-00183767.

PR 09-JUN-2000; 2000JP-00241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saio K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

PS Claim 8; SEQ ID NO 11537; 2537pp + Sequence listing; English.

CC The present invention describes primer sets for synthesizing 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination
 CC of the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 CC
 XX

SQ Sequence 417 AA;

Query Match 17.1%; Score 7; DB 4; Length 417;
 Best Local Similarity 100.0%; Pred. No. 4.9;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 27 LTSHOSQ 33
 |||||
 Db 46 LTSHOSQ 52

Search completed: February 1, 2005, 15:20:12
 Job time : 71.7778 secs

RESULT 15
 AAY49137
 ID AAY49137 standard; protein; 431 AA.
 XX
 AC AAY49137;
 XX
 DT 17-JAN-2000 (first entry)
 XX
 DE Amino acid sequence of choline-binding protein fragment #1.
 XX
 KW Truncated surface binding protein; alpha helix; choline binding protein;
 KW vaccine; invasive bacterial infection; otitis media; sepsis; meningitis;
 KW lobar pneumonia infection; antibody; immature immune system;
 KW immunocompromised.
 XX
 OS Streptococcus pneumoniae.
 XX
 EN WO9951266-A2.
 XX
 PD 14-OCT-1999.
 XX
 PF 06-APR-1999; 99WO-US007680.
 XX
 PR 07-APR-1998; 98US-0080878P.
 PR 15-MAY-1998; 98US-0085743P.
 XX
 PA (MEDI-) MEDIMMUNE INC.
 XX
 PI Wizemann TM, Koenig S, Johnson LS;
 DR WI; 1999-601465/51.
 DR N-PSDB; AA231077.
 XX
 PT New pneumococcal proteins useful as vaccines and for diagnosis of
 PT pneumococcal infections.
 XX
 PS Claim 10; Page 64-65; 98pp; English.
 CC
 CC AAY49137-Y49152 are amino acid sequences that are fragments of choline
 CC binding proteins (CBP). The fragments of the protein are the alpha helix
 CC forming parts of the CBPs from Streptococcus pneumoniae. The polypeptides
 CC do not contain the actual choline binding fragment. The polypeptides and
 CC the nucleotide sequences that encode them (AA231077-231092) are used in
 CC the invention, which relates to polypeptide truncates of a pneumococcal
 CC surface binding protein containing the highly conserved immunogenic alpha
 CC helical portion and no choline binding portion. The polypeptides are used
 CC as immunogens in a bacterial vaccine. The vaccine can be used for
 CC preventing (immunising) or treating invasive bacterial (especially
 CC pneumococcal) infections, especially otitis media (caused by
 CC S.pneumoniae), sepsis, meningitis and lobar pneumonia infections.
 CC Antibodies raised against the polypeptide are useful for detection,
 CC prevention (passive immunity) and treatment of S. pneumoniae infections.
 CC The vaccines are especially useful in immunocompromised patients, those
 CC with an immature immune system, or patients with an on going pneumococcal
 CC infection. The vaccine avoids unnecessary expense and provides broad
 CC protection against a range of pneumococcal serotypes and it produces an
 CC improved and enhanced effect in preventing bacterial infections
 CC
 SQ Sequence 431 AA;

Query March 17.1%; Score 7; DB 2; Length 431;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 KKEELTS 29
 |||||
 Db 74 KKEELTS 80

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:53:01 ; Search time 69.7778 Seconds
(without alignment)
210.782 Million cell updates/sec

Title: SEQ5ASPI967

Perfect score: 41
Sequence: 1 LKGLFTLFAHLVXFPADTL.....QVNISKTDFAFPDSENDPEK 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2002273 seqs, 358729239 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1316031

Minimum DB seq length: 21

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	515	2	AAW54099 Homo sapi
2	20	48.8	349	4	AAW54099 Homo sapi
3	20	48.8	1149	7	ADBO8012 Novel pro
4	20	48.8	2144	4	AAW54099 Homo sapi
5	7	17.1	220	2	AAW54099 Homo sapi
6	7	17.1	453	5	ABW55123 Lactococ
7	7	17.1	587	7	ABO82733 Pseudom
8	7	17.1	1045	2	ABW54099 Homo sapi
9	7	17.1	1045	2	AAW54099 Homo sapi
10	7	17.1	1045	8	ADW98779 HMG-CoA r
11	7	17.1	1045	8	ADW98779 HMG-CoA r
12	7	17.1	1045	8	ADW98779 HMG-CoA r
13	7	17.1	1045	8	ADW98779 HMG-CoA r
14	7	17.1	1045	8	ADW98779 HMG-CoA r
15	6	14.6	45	4	AAW54099 Homo sapi
16	6	14.6	45	4	AAW54099 Homo sapi
17	6	14.6	45	4	AAW54099 Homo sapi
18	6	14.6	45	4	AAW54099 Homo sapi
19	6	14.6	45	4	AAW54099 Homo sapi
20	6	14.6	45	4	AAW54099 Homo sapi
21	6	14.6	45	4	AAW54099 Homo sapi
22	6	14.6	45	4	AAW54099 Homo sapi
23	6	14.6	45	4	AAW54099 Homo sapi
24	6	14.6	45	4	AAW54099 Homo sapi
25	6	14.6	45	4	AAW54099 Homo sapi

26	6	14.6	51	4	ABB44072 Peptide #
27	6	14.6	65	7	ADW57598 Bacterial
28	6	14.6	71	5	ABP03187 Human ORF
29	6	14.6	74	5	ABP02661 Human ORF
30	6	14.6	85	4	ABG02711 Novel hum
31	6	14.6	112	3	AAW54099 Homo sapi
32	6	14.6	112	5	ADW98791 HMG-CoA r
33	6	14.6	115	8	ADW98791 HMG-CoA r
34	6	14.6	115	8	ADW98791 HMG-CoA r
35	6	14.6	115	8	ADW98791 HMG-CoA r
36	6	14.6	118	4	AAW54099 Homo sapi
37	6	14.6	120	4	AAW54099 Homo sapi
38	6	14.6	120	6	ABW71043 Staphyloc
39	6	14.6	121	5	ABP38311 Staphyloc
40	6	14.6	129	3	AAW54099 Homo sapi
41	6	14.6	129	4	AAW54099 Homo sapi
42	6	14.6	133	4	AAW54099 Homo sapi
43	6	14.6	133	4	AAW54099 Homo sapi
44	6	14.6	133	4	AAW54099 Homo sapi
45	6	14.6	144	4	AAW54099 Homo sapi

ALIGNMENTS

RESULT 1	AAW54099	AAW54099 standard; protein; 515 AA.
ID	AAW54099	
XX	AAW54099	
AC	AAW54099	
XX	28-SEP-1998	(first entry)
DT	28-SEP-1998	
XX	Homo sapiens BAP28 sequence.	
DB	Homo sapiens BAP28 sequence.	
XX	BARD1, ring protein; BRCA1; breast cancer; risk; diagnosis.	
KW	Homo sapiens.	
XX	MO9812327-A2	
OS	26-MAR-1998.	
XX	19-SEP-1997;	97WO-US016842.
PF	20-SEP-1996;	96US-0025296P.
PR	03-APR-1997;	97US-0042611P.
PR	04-APR-1997;	97US-0042985P.
XX	(TEKA) UNIV TEXAS SYSTEM.	
PA	Bowcock AM, Baer R;	
XX	WP1; 1998-230317/20.	
DR	N-PSDB; AAV24135.	
XX	DNA sequence encoding BARD1, B123, BE2, BE14, BE31 or BE445 - which as	
PT	breast cancer antigen, BRCA1, binding proteins are useful to identify	
PT	patient having or at risk of developing cancer.	
PS	Disclosure; Page 287-288; 348pp; English.	
XX	The sequence is that of a protein which can be used in the preparation of	
CC	the recombinant breast cancer antigen, BRCA1, binding proteins BARD1,	
CC	B123, BE2, BE14, BE31 or BE445, or a composition for the detection of a	
CC	BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid sequence, specifically	
CC	a wild type BARD1 composition for the detection or purification of BRCA1,	
CC	useful to identify a patient having, or at risk of developing cancer.	
CC	BARD1 can be used in the preparation of an anti-BARD1 antibody, and in	
CC	the detection and purification of a BRCA1 protein. BARD1, B123, BE2,	
CC	BE14, BE31 or BE445 can be used in the identification of a binding protein	
CC	agonist or antagonist that alters the binding of BARD1, B123, BE2, BE14,	
CC	BE31 or BE445 to BRCA1 or the biological activity of the BRCA1-BARD1,	

CC B123, BE2, BE14, BE31 or BE445 complex. The antibodies can be used to
CC detect BARD1, B123, BE2, BE14, BE31 or BE445, a specific anti-BARD1
CC antibody can be used to identify a patient having or at risk of
CC developing cancer
XX
SQ Sequence 515 AA;

Query Match 100.0%; Score 41; DB 2; Length 515;
Best Local Similarity 100.0%; Pred. No. 5,8e-36;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LKGLFTLFGHLLVFPADTLDDQVNIKTDEAFPSNDPEK 41
DB 318 LKGLFTLFGHLLVFPADTLDDQVNIKTDEAFPSNDPEK 358

RESULT 2

AB92729
ID AAB92729 standard; protein; 349 AA.

AC AAB92729;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:11159.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-00116126.

PR 29-JUL-1999; 99JP-00248036.

PR 27-AUG-1999; 99JP-00300253.

PR 11-JAN-2000; 2000JP-00118776.

PR 02-MAY-2000; 2000JP-00183767.

PR 09-JUN-2000; 2000JP-00241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

PS Claim 8; SEQ ID NO 11159; 2537BP + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-
XX length cDNAs defined in the specification. Where a primer set comprises:
XX (a) an oligo-dT primer and an oligonucleotide complementary to the
XX complementary strand of a polynucleotide which comprises one of the 5602
XX nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in the
XX specification. The primer sets can be used in antisense therapy and in
XX gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH1632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX

SQ Sequence 349 AA;

Query Match 48.8%; Score 20; DB 4; Length 349;
Best Local Similarity 100.0%; Pred. No. 2,7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LKGLFTLFGHLLVFPADTL 20
DB 152 LKGLFTLFGHLLVFPADTL 171

RESULT 3

ADE08012
ID ADE08012 standard; protein; 1149 AA.

AC ADE08012;

DT 29-JAN-2004 (first entry)

DE Novel protein (useful for identifying genetic disorders) #167.

KW novel gene; novel protein; tissue marker; molecular weight marker;

KW chromosome marker; genetic disorder.

OS Unidentified.

PN WO2003054152-A2.

PD 03-JUL-2003.

PF 10-DEC-2002; 2002WO-US039555.

PR 10-DEC-2001; 2001US-0339739P.

PR 11-DEC-2001; 2001US-0339653P.

PR 14-MAR-2002; 2002US-0365091P.

PR 14-MAR-2002; 2002US-0365384P.

PR 12-APR-2002; 2002US-0372815P.

PR 22-APR-2002; 2002US-0012855P.

PR 24-APR-2002; 2002US-0376045P.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;

PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Demanac RT, Wang Z;

PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;

DR WPI; 2003-569235/53.

DR N-PSDB; ADE07101.

PS Claim 20; SEQ ID NO 1078; 1177BP; English.

XX The invention comprises the amino acid and coding sequences of novel
XX proteins. The DNA and protein sequences of the invention are useful as:
XX markers for tissues in which the corresponding protein is preferentially
XX expressed; as molecular weight markers on gels; as chromosome markers or
XX tags; to identify chromosome sequences or to map related gene positions; and to
XX compare with endogenous DNA sequences in patients to identify potential
XX genetic disorders. The present amino acid sequence represents a protein
XX of the invention.
XX Sequence 1149 AA;

Query Match 48.8%; Score 20; DB 7; Length 1149;
 Best Local Similarity 100.0%; Pred. No. 7,7e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LKGLFTLFAAGHLVKKFPADTL 20
 |||||
 Db 952 LKGLFTLFAAGHLVKKFPADTL 971

RESULT 4

AAB85029 standard; protein; 2144 AA.

AC AAB85029;

DT 06-AUG-2001 (first entry)

DE Protein encoded by BAP28 cDNA consisting of exons 1 to 45.

XX BAP28; prostate; tumour; cancer; diagnostic; genetic analysis.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 1694

FT /label= Ser or Asn

FT Misc-difference 1854

FT /label= Ala or Val

FT Misc-difference 1967

FT /label= Asp or Asn

FT Misc-difference 2017

FT /label= Gly or Glu

XX WO200100669-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-IB001183.

XX 25-JUN-1999; 99US-0141322P.

XX 18-JAN-2000; 2000US-0176880P.

XX (GERT) GENSET.

PI Barry C, Bougueleret L, Chumakov I, Cohen-Akenine A;

XX WPI; 2001-367032/38.

XX N-PSDB; AAF83909, AAF83910.

XX New BAP28 polynucleotides and polypeptides overexpressed in prostate

XX cancer cells for diagnosing prostate tumors, e.g. by hybridization or

XX polymerase chain reaction assays.

XX Claim 14; Page 297-304; 349pp; English.

XX The invention is directed to BAP28 polypeptides, BAP28 polynucleotide

XX sequences and regulatory region located at the 3' and 5' ends of the

XX BAP28 coding region. The BAP28 polypeptides can be expressed by standard

XX recombinant methodology. BAP28 polynucleotides and polypeptides have been

XX found to be over expressed in prostate tumour cells, therefore levels of

XX BAP28 expression and/or activity may be assayed (e.g. by polymerase chain

XX reaction (PCR)) to diagnose patient suffering from or susceptible to

XX prostate cancer. Antibodies specific for the BAP28 polypeptides are

XX useful as diagnostic reagents. Biallelic markers of the BAP28 gene are

XX useful in genetic analysis. The present sequence represents a protein

XX encoded by a first cDNA sequence of the BAP28 gene consisting of the

XX exons 1 to 45

XX Sequence 2144 AA;

XX Query Match 48.8%; Score 20; DB 4; Length 2144;

XX Best Local Similarity 100.0%; Pred. No. 1.3e-12;

XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LKGLFTLFAAGHLVKKFPADTL 20
 |||||
 Db 1947 LKGLFTLFAAGHLVKKFPADTL 1966

RESULT 5

AAV23793 standard; protein; 220 AA.

AC AAV23793;

DT 27-AUG-2003 (revised)

DT 14-SEP-1999 (first entry)

DE A. gyrase protein of Chitinophaga pinensis.

XX Identification; detection; microbe; gyrase gene; gyrase protein.

XX Chitinophaga pinensis.

XX JP11169175-A.

XX 29-JUN-1999.

XX 12-DEC-1997; 97JP-00343316.

XX 12-DEC-1997; 97JP-00343316.

XX (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.

XX WPI; 1999-422615/36.

XX N-PSDB; AAX86000.

XX Identification and detection of a microbe - by detection of a gyrase

XX gene.

XX Example 3; Page 14-15; 42pp; Japanese.

XX The specification describes a method for the identification or detection

XX of a microbe, using the gyrase gene as the index. The method involves the

XX use of PCR primers to amplify DNA from the microbe, which is then

XX identified or detected depending on its base sequence. The method can be

XX used to classify and identify an unidentified microbe strain rapidly and

XX with high precision. The present sequence represents a gyrase protein.

XX (Updated on 27-AUG-2003 to correct OS field.)

XX Sequence 220 AA;

XX Qy 35 SENDPER 41
 |||||

Db 130 SENDPER 136

RESULT 6

AAB55123 standard; protein; 453 AA.

AC AAB55123;

DT 29-AUG-2003 (revised)

DT 16-MAY-2002 (first entry)

DE Lactococcus lactis protein yedA.

XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.

XX Lactococcus lactis; IL1403.

FN FR2807446-A1.
 XX
 PD 12-OCT-2001.
 XX
 PF 11-APR-2000; 2000FR-00004630.
 XX
 PR 11-APR-2000; 2000FR-00004630.
 XX
 PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
 XX
 PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
 DR WPI; 2002-043418/06.
 XX
 PT New nucleotide sequence useful in the identification or Lactococcus
 PT lactis and related species.
 XX
 PS Claim 6; SEQ ID NO 1825; 2504pp; French.
 CC The present invention is related to a Lactococcus lactis nucleotide
 CC sequence (ABA90521) and related proteins (AB553300-AB555621). The nucleic
 CC acid sequence is useful in the detection and/or amplification of nucleic
 CC acid sequence, particularly to identify Lactococcus lactis or related
 CC species. The proteins of the invention are useful for the biosynthesis or
 CC biodegradation of a composition of interest. The invention helps research
 CC in lactic bacteria, particularly useful in the production of yogurt and
 CC cheese. Note: The sequence data for this patent is based on equivalent
 CC patent WO200177334 (published 18-OCT-2001) which is available in
 CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 29-AUG-2003 to
 CC standardise OS field)
 XX
 SQ Sequence 453 AA;
 Query Match 17.1%; Score 7; DB 5; Length 453;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 22 QVNISKT 26
 Db 127 QVNISKT 133
 RESULT 7
 ABO82733
 ID ABO82733 standard; protein; 587 AA.
 XX
 AC ABO82733;
 XX
 DT 29-JUL-2004 (first entry)
 DE Pseudomonas aeruginosa polypeptide #14908.
 XX
 DE Pseudomonas aeruginosa polypeptide #14908.
 XX
 KW Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.
 XX
 OS Pseudomonas aeruginosa.
 XX
 PN US6551795-B1.
 XX
 PD 22-APR-2003.
 XX
 PF 18-FEB-1999; 99US-00252991.
 XX
 PR 18-FEB-1998; 98US-0074788P.
 PR 27-JUL-1998; 98US-0094190P.
 XX
 PA (GENO-) GENOME THERAPEUTICS CORP.
 XX
 PI Rubenfield MT, Noiling J, Deloughery C, Bush D;
 XX WPI; 2003-615309/58.
 DR N-PSDB; ABD16304.
 XX

PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
 PT useful as molecular targets for diagnostics, prophylaxis and treatment of
 PT pathological conditions resulting from bacterial infection.
 XX
 XX Disclosure; SEQ ID NO 31479; 455pp; English.
 CC
 CC The invention relates to Pseudomonas aeruginosa polypeptides and the
 CC polynucleotides encoding them. The sequences are useful in diagnosis and
 CC therapy of pathological conditions, as molecular targets for diagnosis,
 CC prophylaxis and treatment of pathological conditions resulting from a
 CC bacterial infection, for evaluating a compound, such as a polypeptide,
 CC for the ability to bind a P. aeruginosa nucleic acid, as components of
 CC effective antibacterial targets, as targets for antibacterial drugs,
 CC including anti-P. aeruginosa drugs, as templates for recombinant
 CC production of P. aeruginosa-derived peptides or polypeptides, as target
 CC components for diagnosis and/or treatment of P. aeruginosa-caused
 CC infection, and in detection of P. aeruginosa sequences or other sequences
 CC of Pseudomonas species using biochip technology. Sequences ABO67826-
 CC ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format from USPTO at
 CC seqdata.uspto.gov/sequence.html
 XX
 SQ Sequence 587 AA;
 Query Match 17.1%; Score 7; DB 7; Length 587;
 Best Local Similarity 100.0%; Pred. No. 56;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 17 ADTLDGV 23
 Db 381 ADTLDGV 387
 RESULT 8
 ABP60399
 ID ABP60399 standard; protein; 1045 AA.
 XX
 AC ABP60399;
 XX
 DT 21-MAR-2003 (first entry)
 DE Yeast HMG2 SEQ ID NO 6.
 XX
 DE Yeast HMG2 SEQ ID NO 6.
 XX
 KW Yeast; HMG-CoA reductase; squalene; zymosterol; cholesterol; 7,24-dienol;
 KW ergosta-5, 7, 24-trienol; zymosterol-24-methyl transferase;
 KW ergosta-5, 7, 24 (28)-trienol-22-dehydrogenase; ergosta; dienol;
 KW episterol-5-dehydrogenase; linker region; catalytic domain;
 KW membrane binding region; HMG2; enzyme.
 XX
 OS Saccharomyces cerevisiae.
 XX
 PN US5460949-A.
 XX
 PD 24-OCT-1995.
 XX
 PF 28-OCT-1991; 91US-00783861.
 XX
 PR 15-NOV-1990; 90US-00613380.
 XX
 PA (STAD) AMOCO CORP.
 XX
 PI Mukharji I, Saunders CA, Wolf FR;
 XX WPI; 1992-168867/21.
 DR N-PSDB; ABZ26037.
 XX
 PT Increasing squalene and specific sterol accumulation in yeasts - by
 PT transforming mutant yeasts to increase 3-hydroxy-3-methylglutaryl COA
 PT reductase activity in the yeasts.
 XX
 PS Disclosure; Col 63-68; 60pp; English.
 XX

CC The invention relates to: (A) a method of increasing squalene,
 CC zymosterol, cholesterol-7,24-dienol and cholesterol-5,7,24-trienol
 CC accumulation in yeast comprising increasing the expression level of a
 CC structural gene encoding a polypeptide having HMG-CoA reductase activity
 CC in a mutant yeast having defects in the expression of zymosterol-24-
 CC methyl transferase and ergosta-5,7,24(28)-trienol-22-dehydrogenase;
 CC (B) a method of increasing squalene, ergosta-8,22-dienol, ergosta-7,22-
 CC dieno, ergosta-8-enol and ergosta-7-enol accumulation in *S. cerevisiae*
 CC comprising transforming a mutant *S. cerevisiae* having a defect in the
 CC expression of episterol-5-dehydrogenase with a recombinant DNA molecule
 CC comprising a vector operatively linked to an exogenous DNA segment that
 CC encodes the catalytic region and at least a portion of the linker region
 CC but is free from the membrane binding region of an HMG-CoA reductase
 CC enzyme and a promoter suitable for driving the expression of the
 CC reductase in the yeast; (C) a method of increasing squalene, zymosterol
 CC and cholesterol-7,24-dienol accumulation in *S. cerevisiae* comprising
 CC transforming a mutant *S. cerevisiae* having a defect in the expression of
 CC zymosterol-24-methyl transferase and episterol-5-dehydrogenase with a
 CC recombinant DNA molecule as in (E); (D) a method of increasing squalene,
 CC zymosterol, ergosta-5,7,24(28)-trienol and ergosta-5,7-dienol
 CC accumulation in *S. cerevisiae* comprising transforming a mutant *S.*
 CC *cerevisiae* having a defect in the expression of ergosta-5,7,24(28)-
 CC trienol-22-dehydrogenase with a recombinant DNA molecule as in (B); (E) a
 CC mutant *S. cerevisiae* having defects in the expression of zymosterol-24-
 CC methyl transferase and ergosta-5,7,24(28)-trienol-22-dehydrogenase
 CC enzymes, which mutant is designated ATCO402mm; (F) a mutant of *S.*
 CC *cerevisiae* having single or double defects in the expression of enzymes
 CC that catalyse the conversion of squalene to ergosterol, transformed with
 CC a recombinant DNA molecule as in (B). The present sequence is that of the
 CC *Saccharomyces cerevisiae* HMG-CoA reductase 2 (HGM2) protein of the
 CC invention

SO Sequence 1045 AA;

Query Match 17.1%; Score 7; DB 2; Length 1045;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 HLVPKPA 17

DB 10 HLVPKPA 16

RESULT 9

AAR58611

ID AAR58611 standard; protein, 1045 AA.

AC AAR58611;

XX

DT 25-MAR-2003 (revised)

DT 28-APR-1995 (first entry)

XX

DE Yeast HMG-CoA reductase 2.

XX

XX HMG-CoA reductase 2; 3-hydroxy-3-methylglutaryl coenzyme A; squalene;

KM sterol.

XX

OS *Saccharomyces cerevisiae*.

XX

PN US5349126-A.

XX

PD 20-SEP-1994.

XX

PF 14-AUG-1992; 92US-00934374.

XX

PR 12-OCT-1990; 90US-00596467.

XX

PA (STRAD) AMOCO CORP.

XX

PI Wolf FR, Saunders CA, Chappell J;

XX

DR WPI; 1994-302280/37.

XX

DR N-PSDB; AAQ70611.

XX Transgenic plants with increased insect resistance - by transformation
 PT with DNA encoding HMG-CoA activity to increase squalene and sterol
 PT accumulation.

PS Disclosure; Page 53; 56pp; English.

XX AAQ70611 codes for a yeast HMG-CoA reductase 2, AAR58611. Using the A.
 CC tenebrionis vector pKYLX71, AAQ70611 was used to produce transgenic
 CC plants with increased HMG-CoA reductase activity. Plants with increased
 CC HMG-CoA reductase activity were found to have greater squalene and sterol
 CC accumulation, and therefore increased insect resistance. (Updated on 25-
 CC MAR-2003 to correct PF field.)

XX Sequence 1045 AA;

Query Match 17.1%; Score 7; DB 2; Length 1045;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 HLVPKPA 17

DB 10 HLVPKPA 16

RESULT 10

ADM98779

ID ADM98779 standard; protein, 1045 AA.

XX

AC ADM98779;

XX

DT 01-JUL-2004 (first entry)

XX

DE HMG-CoA reductase polypeptide #32.

XX

XX Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;

KW diterpene synthase; defence toxin; volatile defensive signal;

KW pollinator attractant; photoprotectant; HMG-CoA reductase; enzyme.

XX

OS *Saccharomyces cerevisiae*.

XX

PN US2004072323-A1.

XX

PD 15-APR-2004.

XX

PF 07-JAN-2002; 2002US-00041018.

XX

PR 05-JAN-2001; 2001US-0259880P.

XX

PA (MATSU) MATSUDA S P T.

XX

PA (HART) HART E A.

XX

PI Matsuda SPT, Hart EA;

XX

DR WPI; 2004-373921/35.

XX

PT New unicellular organisms comprising exogenous nucleic acids encoding a

PT geranylgeranyl pyrophosphate and a diterpene synthase, useful for

PT producing diterpene and diterpene precursors.

XX

PS Disclosure; SEQ ID NO 199; 38pp; English.

XX

XX The invention relates to a unicellular organism for producing a diterpene

CC or diterpene precursor comprising an exogenous nucleic acid sequence

CC encoding a geranylgeranyl pyrophosphate synthase under the control of a

CC promoter operable in the organism, and an exogenous nucleic acid sequence

CC encoding a diterpene synthase under the control of a promoter operable in

CC the organism. The invention also relates to methods of producing a

CC diterpene or diterpene precursor and a method of isolating a diterpene

CC synthase comprising growing several cells in the presence of a

CC polyaromatic resin to make a cell/resin mixture, where at least one of

CC the cells further comprises at least one isolated and purified nucleic

CC acid sequence of a yeast expression library, and the expression of the

CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents an HMG-CoA reductase polypeptide used in the scope of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification but was obtained in electronic format from
CC USPTO at seqdata.uspto.gov/sequence.html.
CC XX

SO Sequence 1045 AA;

Query Match 17.1%; Score 7; DB 8; Length 1045;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 HLVPFA 17
Db 10 HLVPFA 16

RESULT 11
ADM98798
ID ADM98798 standard; protein; 1045 AA.
XX
AC ADM98798;
XX
DT 01-JUL-2004 (first entry)
XX
DE HMG-CoA reductase polypeptide #51.
XX
KW Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;
KW diterpene synthase; defence toxin; volatile defensive signal;
KW pollinator attractant; photoprotectant; HMG-CoA reductase; enzyme.
XX
OS Saccharomyces cerevisiae.
XX
FN US2004072323-A1.
XX
PD 15-APR-2004.
XX
PE 07-JAN-2002; 2002US-00041018.
XX
PR 05-JAN-2001; 2001US-0259880P.
XX
PA (MATSU) MATSUDA S P T.
PA (HART) HART E A.
XX
PI Matsuda SPT, Hart EA;
XX
DR WPI; 2004-373921/35.
XX
PT New unicellular organisms comprising exogenous nucleic acids encoding a
PT geranylgeranyl pyrophosphate and a diterpene synthase, useful for
PT producing diterpenes and diterpene precursors.
XX
PS Disclosure; SEQ ID NO 218; 38pp; English.
XX

CC The invention relates to a unicellular organism for producing a diterpene
CC or diterpene precursor comprising an exogenous nucleic acid sequence
CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
CC promoter operable in the organism, and an exogenous nucleic acid sequence
CC encoding a diterpene synthase under the control of a promoter operable in
CC the organism. The invention also relates to methods of producing a
CC diterpene or diterpene precursor and a method of isolating a diterpene
CC synthase comprising growing several cells in the presence of a
CC polyaromatic resin to make a cell/resin mixture, where at least one of
CC the cells further comprises at least one isolated and purified nucleic
CC acid sequence of a yeast expression library, and the expression of the

CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents an HMG-CoA reductase polypeptide used in the scope of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification but was obtained in electronic format from
CC USPTO at seqdata.uspto.gov/sequence.html.
CC XX

SO Sequence 1045 AA;

Query Match 17.1%; Score 7; DB 8; Length 1045;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 HLVPFA 17
Db 10 HLVPFA 16

RESULT 12
ADM98844
ID ADM98844 standard; protein; 1045 AA.
XX
AC ADM98844;
XX
DT 01-JUL-2004 (first entry)
XX
DE HMG-CoA reductase polypeptide #97.
XX
KW Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;
KW diterpene synthase; defence toxin; volatile defensive signal;
KW pollinator attractant; photoprotectant; HMG-CoA reductase; enzyme.
XX
OS Saccharomyces cerevisiae.
XX
FN US2004072323-A1.
XX
PD 15-APR-2004.
XX
PE 07-JAN-2002; 2002US-00041018.
XX
PR 05-JAN-2001; 2001US-0259880P.
XX
PA (MATSU) MATSUDA S P T.
PA (HART) HART E A.
XX
PI Matsuda SPT, Hart EA;
XX
DR WPI; 2004-373921/35.
XX
PT New unicellular organisms comprising exogenous nucleic acids encoding a
PT geranylgeranyl pyrophosphate and a diterpene synthase, useful for
PT producing diterpenes and diterpene precursors.
XX
PS Disclosure; SEQ ID NO 264; 38pp; English.
XX

CC The invention relates to a unicellular organism for producing a diterpene
CC or diterpene precursor comprising an exogenous nucleic acid sequence
CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
CC promoter operable in the organism, and an exogenous nucleic acid sequence
CC encoding a diterpene synthase under the control of a promoter operable in
CC the organism. The invention also relates to methods of producing a
CC diterpene or diterpene precursor and a method of isolating a diterpene
CC synthase comprising growing several cells in the presence of a
CC polyaromatic resin to make a cell/resin mixture, where at least one of
CC the cells further comprises at least one isolated and purified nucleic
CC acid sequence of a yeast expression library, and the expression of the

CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents an HMG-CoA reductase polypeptide used in the scope of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification but was obtained in electronic format from
CC USPTO at seqdata.uspto.gov/sequence.html.
CC XX

XX Sequence 1045 AA;

Query Match 17.1%; Score 7; DB 8; Length 1045;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 HLKRPFA 17
|||
Db 10 HLKRPFA 16

RESULT 13

ADM98916 ID ADM98916 standard; protein; 1045 AA.

AC ADM98916;

DT 01-JUL-2004 (first entry)

XX HMG-CoA reductase polypeptide #169.

XX Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;

KW diterpene synthase; defence toxin; volatile defensive signal;

KW pollinator attractant; photoprotectant; HMG-CoA reductase; enzyme.

OS Saccharomyces cerevisiae.

XX US2004072323-A1.

XX 15-APR-2004.

XX 07-JAN-2002; 2002US-00041018.

XX 05-JAN-2001; 2001US-0259880P.

XX (MATS/) MATSUDA S P T.

XX (HART/) HART E A.

XX Matsuda SPT, Hart EA;

XX WPI; 2004-373921/35.

XX New unicellular organisms comprising exogenous nucleic acids encoding a

XX geranylgeranyl pyrophosphate and a diterpene synthase, useful for

XX producing diterpenes and diterpene precursors.

XX Disclosure; SEQ ID NO 336; 38pp; English.

XX The invention relates to a unicellular organism for producing a diterpene

XX or diterpene precursor comprising an exogenous nucleic acid sequence

XX encoding a geranylgeranyl pyrophosphate synthase under the control of a

XX promoter operable in the organism, and an exogenous nucleic acid sequence

XX encoding a diterpene synthase under the control of a promoter operable in

XX the organism. The invention also relates to methods of producing a

XX diterpene or diterpene precursor and a method of isolating a diterpene

XX synthase comprising growing several cells in the presence of a

XX polyaromatic resin to make a cell/resin mixture, where at least one of

XX the cells further comprises at least one isolated and purified nucleic

XX acid sequence of a yeast expression library, and the expression of the

CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents an HMG-CoA reductase polypeptide used in the scope of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification but was obtained in electronic format from
CC USPTO at seqdata.uspto.gov/sequence.html.
CC XX

XX Sequence 1045 AA;

Query Match 17.1%; Score 7; DB 8; Length 1045;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 HLKRPFA 17
|||
Db 10 HLKRPFA 16

RESULT 14

ADM98889 ID ADM98889 standard; protein; 1045 AA.

AC ADM98889;

DT 01-JUL-2004 (first entry)

XX HMG-CoA reductase polypeptide #142.

XX Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;

KW diterpene synthase; defence toxin; volatile defensive signal;

KW pollinator attractant; photoprotectant; HMG-CoA reductase; enzyme.

OS Saccharomyces cerevisiae.

XX US2004072323-A1.

XX 15-APR-2004.

XX 07-JAN-2002; 2002US-00041018.

XX 05-JAN-2001; 2001US-0259880P.

XX (MATS/) MATSUDA S P T.

XX (HART/) HART E A.

XX Matsuda SPT, Hart EA;

XX WPI; 2004-373921/35.

XX New unicellular organisms comprising exogenous nucleic acids encoding a

XX geranylgeranyl pyrophosphate and a diterpene synthase, useful for

XX producing diterpenes and diterpene precursors.

XX Disclosure; SEQ ID NO 309; 38pp; English.

XX The invention relates to a unicellular organism for producing a diterpene

XX or diterpene precursor comprising an exogenous nucleic acid sequence

XX encoding a geranylgeranyl pyrophosphate synthase under the control of a

XX promoter operable in the organism, and an exogenous nucleic acid sequence

XX encoding a diterpene synthase under the control of a promoter operable in

XX the organism. The invention also relates to methods of producing a

XX diterpene or diterpene precursor and a method of isolating a diterpene

XX synthase comprising growing several cells in the presence of a

XX polyaromatic resin to make a cell/resin mixture, where at least one of

XX the cells further comprises at least one isolated and purified nucleic

XX acid sequence of a yeast expression library, and the expression of the

CC nucleic acid sequence is regulated by an inducible promoter under
 CC conditions where the expression is induced, filtering the cell/resin
 CC mixture, extracting the cell/resin mixture with alcohol to produce an
 CC organic eluent and analysing the organic eluent by a screening method
 CC including chromatography and/or spectroscopy, to identify the nucleic
 CC acid sequence encoding the diterpene synthase. The unicellular
 CC microorganism is useful as a diterpene or diterpene precursor producing
 CC system. Diterpenes, in plants, serve as defence toxins, volatile
 CC defensive signals, pollinator attractants and photoprotectants. This
 CC sequence represents an HMG-CoA reductase polypeptide used in the scope of
 CC the invention. Note: The sequence data for this patent did not form part
 CC of the printed specification but was obtained in electronic format from
 CC USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 1045 AA;

Query Match 17.1%; Score 7; DB 8; Length 1045;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 HLVKPPA 17
 |||||
 Db 10 HLVKPPA 16

RESULT 15

AA014903
 ID AA014903 standard; protein; 45 AA.

XX AA014903;

DT 12-OCT-2001 (first entry)

DE Peptide #1337 encoded by probe for measuring cervical gene expression.

XX Probe; human; microarray; gene expression; cervical epithelial cell;

KW cervical cancer.

XX Homo sapiens.

XX WO200157278-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000670.

XX 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human cervical epithelial cells.

XX Claim 27; SEQ ID NO 19729; 487bp; English.

CC The present invention relates to human single exon nucleic acid probes
 CC (SENP; see A0110068-A0128459). The present sequence is a peptide encoded
 CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer. Note: The sequence data for this patent did not form

CC part of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 45 AA;

Query Match 14.6%; Score 6; DB 4; Length 45;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 DTLDQV 23
 |||||
 Db 9 DTLDQV 14

Search completed: February 1, 2005, 15:20:16.
 Job time : 72.7778 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:58:28 ; Search time 17.5556 Seconds
(without alignment)
154.882 Million cell updates/sec

Title: SEQ5ASN1694

Perfect score: 41
Sequence: 1 LKLLCKNFGAENPDPPVPLV.....TAVKLIAPEKKEKVLGSA 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size : 0

Total number of hits satisfying chosen parameters: 230433

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA.*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PTCUTS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	17.1	160	4	US-09-270-767-33214
2	7	17.1	160	4	US-09-270-767-48431
3	6	14.6	51	4	US-09-513-999C-5558
4	6	14.6	61	4	US-09-583-110-4172
5	6	14.6	81	4	US-09-621-976-6450
6	6	14.6	87	4	US-09-513-999C-4616
7	6	14.6	88	4	US-09-248-796A-25228
8	6	14.6	98	4	US-09-540-236-3312
9	6	14.6	99	4	US-09-270-767-35660
10	6	14.6	99	4	US-09-270-767-50877
11	6	14.6	127	4	US-09-270-767-18390
12	6	14.6	127	4	US-09-270-767-53607
13	6	14.6	129	6	5229115-2
14	6	14.6	135	4	US-09-248-796A-22195
15	6	14.6	147	4	US-09-710-279-1214
16	6	14.6	154	1	US-08-446-908-2
17	6	14.6	154	1	US-08-231-205A-2
18	6	14.6	154	2	US-08-871-161-2
19	6	14.6	169	4	US-09-248-796A-19969
20	6	14.6	199	3	US-08-737-248-23
21	6	14.6	205	4	US-09-252-991A-26704
22	6	14.6	218	4	US-09-710-279-450
23	6	14.6	226	4	US-09-248-796A-18711
24	6	14.6	237	4	US-09-248-796A-18839
25	6	14.6	249	4	US-09-248-796A-14685
26	6	14.6	253	2	US-08-685-992-10
27	6	14.6	253	2	US-09-144-925-10

28	6	14.6	294	4	US-09-540-236-3779	Sequence 3779, Ap
29	6	14.6	298	4	US-09-134-000C-5845	Sequence 5845, Ap
30	6	14.6	309	4	US-09-489-039A-8203	Sequence 8203, Ap
31	6	14.6	325	3	US-09-134-001C-3551	Sequence 3551, Ap
32	6	14.6	329	4	US-09-149-476-483	Sequence 483, App
33	6	14.6	339	4	US-09-107-532A-5514	Sequence 5514, App
34	6	14.6	344	4	US-09-248-796A-16383	Sequence 16383, A
35	6	14.6	348	1	US-08-454-196-8	Sequence 8, Appli
36	6	14.6	348	1	US-08-454-196-17	Sequence 17, Appli
37	6	14.6	348	3	US-09-054-033-8	Sequence 8, Appli
38	6	14.6	348	3	US-09-054-033-17	Sequence 17, Appli
39	6	14.6	348	4	US-09-291-046-8	Sequence 8, Appli
40	6	14.6	348	4	US-09-291-046-17	Sequence 17, Appli
41	6	14.6	351	4	US-09-248-796A-16176	Sequence 16176, A
42	6	14.6	372	2	US-08-683-262B-64	Sequence 64, Appli
43	6	14.6	372	3	US-09-361-707-64	Sequence 64, Appli
44	6	14.6	377	3	US-09-150-133-5	Sequence 5, Appli
45	6	14.6	377	3	US-09-150-141-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-270-767-33214
Sequence 33214, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 33214
LENGTH: 160
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-270-767-33214

6mer

Query Match 17.1%; Score 7; DB 4; Length 160;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 VPVINTA 23
Db 52 VPVINTA 58

RESULT 2
US-09-270-767-48431
Sequence 48431, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 48431
LENGTH: 160
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-270-767-48431

Query Match 17.1%; Score 7; DB 4; Length 160;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 17 VPVINTA 23

Db 52 VPVLNTA 58

RESULT 3

US-09-513-999C-5558
 ; Sequence 5558, Application US/09513999C
 ; Patent No. 6783961
 ; GENERAL INFORMATION:
 ; APPLICANT: Dumas Milne Edwards, J.B.
 ; APPLICANT: Duclet, A.Y.
 ; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
 ; Patent No. 6783961
 ; FILE REFERENCE: 59.US2.REG
 ; CURRENT APPLICATION NUMBER: US/09/513,999C
 ; CURRENT FILING DATE: 2000-02-24
 ; PRIOR APPLICATION NUMBER: US 60/122,487
 ; PRIOR FILING DATE: 1999-02-26
 ; NUMBER OF SEQ ID NOS: 36681
 ; SOFTWARE: Patent.pm
 ; SEQ ID NO 5558
 ; LENGTH: 51
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-513-999C-5558

Query Match 14.6%; Score 6; DB 4; Length 51;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 EKKNTL 38
 Db 25 EKKNTL 30

RESULT 4

US-09-583-110-4172
 ; Sequence 4172, Application US/09583110
 ; Patent No. 6699703
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al.
 ; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
 ; FILE REFERENCE: PAT00-07A
 ; CURRENT APPLICATION NUMBER: US/09/583,110
 ; CURRENT FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/107,433
 ; PRIOR FILING DATE: 1998-06-30
 ; PRIOR APPLICATION NUMBER: US 60/085,131
 ; PRIOR FILING DATE: 1998-05-12
 ; PRIOR APPLICATION NUMBER: US 60/051,553
 ; PRIOR FILING DATE: 1997-07-02
 ; NUMBER OF SEQ ID NOS: 5322
 ; SEQ ID NO 4172
 ; LENGTH: 61
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 ; US-09-583-110-4172

Query Match 14.6%; Score 6; DB 4; Length 61;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LCKNPF 8
 Db 11 LCKNPF 16

RESULT 5

US-09-621-976-6450
 ; Sequence 6450, Application US/09621976
 ; Patent No. 6639063

; GENERAL INFORMATION:
 ; APPLICANT: Dumas Milne Edwards, J.B.
 ; APPLICANT: Jobert, S.
 ; APPLICANT: Giordano, J.Y.
 ; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
 ; FILE REFERENCE: GENSET.054PR2
 ; CURRENT APPLICATION NUMBER: US/09/621,976
 ; CURRENT FILING DATE: 2000-07-21
 ; NUMBER OF SEQ ID NOS: 19335
 ; SOFTWARE: Patent.pm
 ; SEQ ID NO 6450
 ; LENGTH: 81
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-621-976-6450

Query Match 14.6%; Score 6; DB 4; Length 81;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 ERKEEK 35
 Db 51 ERKEEK 56

RESULT 6

US-09-513-999C-4616
 ; Sequence 4616, Application US/09513999C
 ; Patent No. 6783961
 ; GENERAL INFORMATION:
 ; APPLICANT: Dumas Milne Edwards, J.B.
 ; APPLICANT: Duclet, A.Y.
 ; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
 ; Patent No. 6783961
 ; FILE REFERENCE: 59.US2.REG
 ; CURRENT APPLICATION NUMBER: US/09/513,999C
 ; CURRENT FILING DATE: 2000-02-24
 ; PRIOR APPLICATION NUMBER: US 60/122,487
 ; PRIOR FILING DATE: 1999-02-26
 ; NUMBER OF SEQ ID NOS: 36681
 ; SOFTWARE: Patent.pm
 ; SEQ ID NO 4616
 ; LENGTH: 87
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SIGNAL
 ; LOCATION: -48...-1
 ; OTHER INFORMATION: score 3.8
 ; OTHER INFORMATION: seq NSLLLLCLVYIP/HS
 ; FEATURE:
 ; NAME/KEY: UNSURE
 ; LOCATION: 7
 ; OTHER INFORMATION: Xaa=asp or Glu
 ; FEATURE:
 ; NAME/KEY: UNSURE
 ; LOCATION: 8
 ; OTHER INFORMATION: Xaa= * or Cys or Phe or Leu or Trp or Tyr
 ; FEATURE:
 ; NAME/KEY: UNSURE
 ; LOCATION: 21
 ; OTHER INFORMATION: Xaa=Ala or Asp or Glu or Gly or Ile or Lys or Met or Asn or Arg o;
 ; FEATURE:
 ; NAME/KEY: UNSURE
 ; LOCATION: 22
 ; OTHER INFORMATION: Xaa=Cys or Phe
 ; US-09-513-999C-4616

Query Match 14.6%; Score 6; DB 4; Length 87;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 LCKNF 8
 Db 15 LCKNF 20

RESULT 7
 US-09-248-796A-25228

Sequence 25228, Application US/09248796A
 Patent No. 6747137
 GENERAL INFORMATION:

APPLICANT: Keith Weinstock et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
 FILE REFERENCE: 107196.132
 CURRENT APPLICATION NUMBER: US/09/248,796A
 PRIOR FILING DATE: 1999-02-12
 PRIOR APPLICATION NUMBER: US 60/074,725
 PRIOR FILING DATE: 1998-02-13
 PRIOR APPLICATION NUMBER: US 60/096,409
 PRIOR FILING DATE: 1998-08-13
 NUMBER OF SEQ ID NOS: 28208
 SEQ ID NO 25228
 LENGTH: 88
 TYPE: PRT
 ORGANISM: Candida albicans

US-09-248-796A-25228

Query Match 14.6%; Score 6; DB 4; Length 88;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 30 ERKEEK 35
 Db 35 ERKEEK 40

RESULT 8

US-09-540-236-3312
 Sequence 3312, Application US/09540236
 Patent No. 6673910
 GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
 FILE REFERENCE: 2709 2005-001
 CURRENT APPLICATION NUMBER: US/09/540,236
 CURRENT FILING DATE: 2000-04-04
 NUMBER OF SEQ ID NOS: 3840
 SEQ ID NO 3312
 LENGTH: 98
 TYPE: PRT
 ORGANISM: M.catarhalis
 US-09-540-236-3312

Query Match 14.6%; Score 6; DB 4; Length 98;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 16 FVPVIL 21
 Db 18 FVPVIL 23

RESULT 9

US-09-270-767-35660
 Sequence 35660, Application US/09270767
 Patent No. 6703491
 GENERAL INFORMATION:

APPLICANT: Homburger et al.
 TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 FILE REFERENCE: File Reference: 7326-094
 CURRENT APPLICATION NUMBER: US/09/270,767
 CURRENT FILING DATE: 1999-03-17

NUMBER OF SEQ ID NOS: 62517
 SOFTWARE: Patentln Ver. 2.0
 SEQ ID NO 35660
 LENGTH: 99
 TYPE: PRT
 ORGANISM: Drosophila melanogaster

FEATURE:
 OTHER INFORMATION: Xaa means any amino acid
 US-09-270-767-35660

Query Match 14.6%; Score 6; DB 4; Length 99;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 13 PDPEVP 18
 Db 86 PDPEVP 91

RESULT 10

US-09-270-767-50877
 Sequence 50877, Application US/09270767
 Patent No. 6703491
 GENERAL INFORMATION:

APPLICANT: Homburger et al.
 TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 FILE REFERENCE: File Reference: 7326-094
 CURRENT APPLICATION NUMBER: US/09/270,767
 CURRENT FILING DATE: 1999-03-17
 NUMBER OF SEQ ID NOS: 62517
 SOFTWARE: Patentln Ver. 2.0
 SEQ ID NO 50877
 LENGTH: 99
 TYPE: PRT
 ORGANISM: Drosophila melanogaster
 FEATURE:
 OTHER INFORMATION: Xaa means any amino acid
 US-09-270-767-50877

Query Match 14.6%; Score 6; DB 4; Length 99;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 13 PDPEVP 18
 Db 86 PDPEVP 91

RESULT 11

US-09-270-767-38390
 Sequence 38390, Application US/09270767
 Patent No. 6703491
 GENERAL INFORMATION:

APPLICANT: Homburger et al.
 TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 FILE REFERENCE: File Reference: 7326-094
 CURRENT APPLICATION NUMBER: US/09/270,767
 CURRENT FILING DATE: 1999-03-17
 NUMBER OF SEQ ID NOS: 62517
 SOFTWARE: Patentln Ver. 2.0
 SEQ ID NO 38390
 LENGTH: 127
 TYPE: PRT
 ORGANISM: Drosophila melanogaster
 FEATURE:
 OTHER INFORMATION: Xaa means any amino acid
 US-09-270-767-38390

Query Match 14.6%; Score 6; DB 4; Length 127;
 Best Local Similarity 100.0%; Pred. No. 76;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 30 ERKEEK 35

Db 107 ERKEEK 112

RESULT 12
US-09-270-767-53607
Sequence 53607, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 53607
LENGTH: 127
TYPE: PRT
ORGANISM: Drosophila melanogaster
FEATURE:
OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-53607

Query Match 14.6%; Score 6; DB 4; Length 127;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 ERKEEK 35
Db 107 ERKEEK 112

RESULT 13
5229115-2
Patent No. 5229115
APPLICANT: LYNCH, DAVID H.
TITLE OF INVENTION: ADOPTIVE IMMUNOTHERAPY WITH INTERLEUKIN-7
NUMBER OF SEQUENCES: 2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/559,001
FILING DATE: 26-JUL-1990
SEQ ID NO: 2
LENGTH: 129
5229115-2

Query Match 14.6%; Score 6; DB 6; Length 129;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 KEKRV 37
Db 94 KEKRV 99

RESULT 14
US-09-248-796A-22195
Sequence 22195, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 22195
LENGTH: 135

TYPE: PRT
ORGANISM: Candida albicans
US-09-248-796A-22195

Query Match 14.6%; Score 6; DB 4; Length 135;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 KNVIGS 40
Db 55 KNVIGS 60

RESULT 15
US-09-710-279-1214
Sequence 1214, Application US/09710279
Patent No. 6703492
GENERAL INFORMATION:
APPLICANT: KIMMERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: P03480US
CURRENT APPLICATION NUMBER: US/09/710,279
CURRENT FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4472
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1214
LENGTH: 147
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-710-279-1214

Query Match 14.6%; Score 6; DB 4; Length 147;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 VKLIAP 29
Db 120 VKLIAP 125

Search completed: February 1, 2005, 15:35:42
Job time : 17.5556 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 15:09:39 ; Search time 59.7778 Seconds
(without alignments)
247.799 Million cell updates/sec

Title: SEQ5ASN1967

Sequence: 1 LKGLFTLFAGLVAKVPADTL.....QVNIKTDFAFDSNDPEK 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1608061 seqs, 361289386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1311380

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubppa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep:*
- 4: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pep:*
- 5: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
- 6: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubppa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
- 17: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep:*
- 19: /cgn2_6/ptodata/1/pubppa/US60_NEW_PUB.pep:*
- 20: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	41	1149	US-10-128-558-167	Sequence 167, App
2	7	17.1	182 16 US-10-437-863-193489	Sequence 193489, App
3	7	17.1	220 9 US-09-823-829-6	Sequence 6, Appl1
4	7	17.1	220 9 US-09-823-823-6	Sequence 6, Appl1
5	7	17.1	1045 14 US-10-369-493-1836	Sequence 1836, App
6	7	17.1	1045 15 US-10-041-018-199	Sequence 199, App
7	7	17.1	1045 15 US-10-041-018-218	Sequence 218, App
8	7	17.1	1045 15 US-10-041-018-264	Sequence 264, App
9	7	17.1	1045 15 US-10-041-018-309	Sequence 309, App
10	7	17.1	1045 15 US-10-041-018-336	Sequence 336, App
11	6	14.6	41 9 US-09-879-957-79	Sequence 79, Appl
12	6	14.6	41 16 US-10-807-856-79	Sequence 79, Appl
13	6	14.6	51 9 US-09-864-761-47882	Sequence 47882, A

14	6	14.6	55	15	US-10-424-599-284582	Sequence 284582, A
15	6	14.6	69	16	US-10-767-701-53471	Sequence 53471, A
16	6	14.6	80	15	US-10-424-599-143186	Sequence 143186, A
17	6	14.6	80	17	US-10-425-115-237591	Sequence 237591, A
18	6	14.6	87	17	US-10-425-115-250582	Sequence 250582, A
19	6	14.6	91	15	US-10-424-599-26946	Sequence 26946, A
20	6	14.6	93	15	US-10-424-599-1482753	Sequence 1482753, A
21	6	14.6	95	16	US-10-767-701-46828	Sequence 46828, A
22	6	14.6	105	15	US-10-424-599-157320	Sequence 157320, A
23	6	14.6	111	15	US-10-424-599-244159	Sequence 244159, A
24	6	14.6	111	17	US-10-425-115-340045	Sequence 340045, A
25	6	14.6	115	15	US-10-041-018-211	Sequence 211, App
26	6	14.6	115	15	US-10-041-018-351	Sequence 351, App
27	6	14.6	118	15	US-10-424-599-177366	Sequence 177366, A
28	6	14.6	125	16	US-10-437-963-114015	Sequence 114015, A
29	6	14.6	127	16	US-10-767-701-56103	Sequence 56103, A
30	6	14.6	129	17	US-10-739-956-10	Sequence 10, Appl
31	6	14.6	133	17	US-10-425-115-194350	Sequence 194350, A
32	6	14.6	143	17	US-10-425-115-342272	Sequence 342272, A
33	6	14.6	146	15	US-10-424-599-174950	Sequence 174950, A
34	6	14.6	150	17	US-10-425-115-351501	Sequence 351501, A
35	6	14.6	171	15	US-10-424-599-148350	Sequence 148350, A
36	6	14.6	187	15	US-10-425-114-43393	Sequence 43393, A
37	6	14.6	189	17	US-10-425-115-200953	Sequence 200953, A
38	6	14.6	190	14	US-10-056-884-6	Sequence 6, Appl1
39	6	14.6	200	14	US-10-739-930-6571	Sequence 6571, App
40	6	14.6	201	17	US-10-424-599-249118	Sequence 249118, A
41	6	14.6	204	15	US-10-767-701-44196	Sequence 44196, A
42	6	14.6	205	16	US-10-425-114-67793	Sequence 67793, A
43	6	14.6	210	15	US-10-788-792-165	Sequence 165, App
44	6	14.6	214	17	US-10-723-860-2757	Sequence 2757, App
45	6	14.6	214	17	US-10-723-860-2757	Sequence 2757, App

ALIGNMENTS

RESULT 1
US-10-128-558-167
Sequence 167, Application US/10128558
Publication No. US20040219521A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Wang, Zhilwei
APPLICANT: Beng, Gezhi
APPLICANT: Boyle, Bryan J
APPLICANT: Drmanac, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
FILE REFERENCE: 812A
CURRENT APPLICATION NUMBER: US/10/128,558
CURRENT FILING DATE: 2002-04-22
PRIOR APPLICATION NUMBER: US 60/339,453
PRIOR FILING DATE: 2001-12-11
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02823
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: PCT/US01/03800
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 09/515,126
PRIOR FILING DATE: 2000-02-28
Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 412
SOFTWARE: pt_fl_genes Version 6.0
SEQ ID NO 167
LENGTH: 1149
TYPE: PRT
ORGANISM: Homo sapiens
US-10-128-558-167

Query Match 100.0%; Score 41; DB 17; Length 1149;
Best Local Similarity 100.0%; Pred. No. 17e-33;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LKGLFTLFGHIVKPPADTLNQVNIKSTDEAFPSSENDEK 41
DB 952 LKGLFTLFGHIVKPPADTLNQVNIKSTDEAFPSSENDEK 992

RESULT 2

US-10-437-963-193489
Sequence 193489, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 193489
LENGTH: 182
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(182)
OTHER INFORMATION: unsure at all Xaa locations
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_89622C.1.pep
US-10-437-963-193489

Query Match 17.1%; Score 7; DB 16; Length 182;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LKGLFTL 7
DB 19 LKGLFTL 25

RESULT 3

US-09-823-829-6
Sequence 6, Application US/09823829
Patent No. US2002014667A1
GENERAL INFORMATION:
APPLICANT: Yamamoto, Satoshi
APPLICANT: Nakamura, Shoko
APPLICANT: Suzuki, Makoto
APPLICANT: Kasai, Hiroaki
APPLICANT: Hamada, Tohtu
TITLE OF INVENTION: METHOD FOR IDENTIFICATION AND DETECTION OF MICROORGANISMS
TITLE OF INVENTION: USING GYRASE GENE AS AN INDICATOR
FILE REFERENCE: 12817-004001
CURRENT APPLICATION NUMBER: US/09/823,829
CURRENT FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: US 09/208,688

PRIOR FILING DATE: 1998-12-10
PRIOR APPLICATION NUMBER: JP 97/343316
PRIOR FILING DATE: 1997-12-12
NUMBER OF SEQ ID NOS: 82
SOFTWARE: PatentIn version 2.0
SEQ ID NO 6
LENGTH: 220
TYPE: PRT
ORGANISM: Chitinophaga pinensis
US-09-823-829-6

Query Match 17.1%; Score 7; DB 9; Length 220;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 SENDPEK 41
DB 130 SENDPEK 136

RESULT 4

US-09-823-823-6
Sequence 6, Application US/09823823
Patent No. US20020171092A1
GENERAL INFORMATION:
APPLICANT: Yamamoto, Satoshi
APPLICANT: Kasai, Hiroaki
APPLICANT: Nakamura, Shoko
APPLICANT: Suzuki, Makoto
APPLICANT: Hamada, Tohtu
TITLE OF INVENTION: METHOD FOR IDENTIFICATION AND DETECTION OF MICROORGANISMS USING G
TITLE OF INVENTION: GENE AS AN INDICATOR
FILE REFERENCE: 12817-004001
CURRENT APPLICATION NUMBER: US/09/823,823
CURRENT FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: US 09/208,688
PRIOR FILING DATE: 1998-12-10
PRIOR APPLICATION NUMBER: JP 97/343316
PRIOR FILING DATE: 1997-12-12
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn version 2.0
SEQ ID NO 6
LENGTH: 220
TYPE: PRT
ORGANISM: Chitinophaga pinensis
US-09-823-823-6

Query Match 17.1%; Score 7; DB 9; Length 220;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 SENDPEK 41
DB 130 SENDPEK 136

RESULT 5

US-10-369-493-1836
Sequence 1836, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1836
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-1836

Query Match 17.1%; Score 7; DB 14; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 HLKVPFA 17
Db 10 HLKVPFA 16

RESULT 6
US-10-041-018-199
; Sequence 199, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041.018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 199
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-041-018-199

Query Match 17.1%; Score 7; DB 15; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 HLKVPFA 17
Db 10 HLKVPFA 16

RESULT 7
US-10-041-018-218
; Sequence 218, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041.018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 218
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-041-018-218

Query Match 17.1%; Score 7; DB 15; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 HLKVPFA 17
Db 10 HLKVPFA 16

Db 10 HLKVPFA 16

RESULT 8
US-10-041-018-264
; Sequence 264, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041.018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 264
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-041-018-264

Query Match 17.1%; Score 7; DB 15; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 HLKVPFA 17
Db 10 HLKVPFA 16

RESULT 9
US-10-041-018-309
; Sequence 309, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041.018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 309
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-041-018-309

Query Match 17.1%; Score 7; DB 15; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 HLKVPFA 17
Db 10 HLKVPFA 16

RESULT 10
US-10-041-018-336
; Sequence 336, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041.018

CURRENT FILING DATE: 2002-01-07
PRIOR APPLICATION NUMBER: US 60/259880
PRIOR FILING DATE: 2001-01-05
NUMBER OF SEQ ID NOS: 413
SOFTWARE: Patentin version 3.1
SEQ ID NO 336
LENGTH: 1045
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-10-041-018-336

Query Match 17.1%; Score 7; DB 15; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 HLVKPEA 17
Db 10 HLVKPEA 16

RESULT 11
US-09-879-957-79
Sequence 79, Application US/09879957
Patent No. US20020034755A1
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
HOFFMAN, No. US20020034755A1h
KAY, Brian K.
FOWLES, Dana M.
MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 41 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-09-879-957-79
Query Match 14.6%; Score 6; DB 9; Length 41;
Best Local Similarity 100.0%; Pred. No. 91;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 36 ENDEPK 41
Db 29 ENDEPK 34

RESULT 12
US-10-807-856-79
Sequence 79, Application US/10807856
Publication No. US20040157216A1
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
HOFFMAN, Noah
KAY, Brian K.
FOWLES, Dana M.
MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/807,856
FILING DATE: 23-Mar-2004
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915
FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 41 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-807-856-79
Query Match 14.6%; Score 6; DB 16; Length 41;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 36 ENDEPK 41
Db 29 ENDEPK 34
RESULT 13
US-09-864-761-47882
Sequence 47882, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.

```

; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wenheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 47882
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008897.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 1.7
; OTHER INFORMATION: EST HUMAN HIT: AUI32001.1, EVALUATE 7.00e-23
; OTHER INFORMATION: SWISSPROT HIT: P04035, EVALUATE 6.00e-24
US-09-864-761-47882

Query Match      14.6%; Score 6; DB 9; Length 51;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 AGHLVK 14
      |||||
DB      40 AGHLVK 45

RESULT 14
US-10-424-599-284582
; Sequence 284582, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
```

```

; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 284582
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_99002C.1.pep
US-10-424-599-284582

Query Match      14.6%; Score 6; DB 15; Length 55;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      11 HLVKPF 16
      |||||
DB      30 HLVKPF 35

RESULT 15
US-10-767-701-53471
; Sequence 53471, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 53471
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 13392228.pep
US-10-767-701-53471

Query Match      14.6%; Score 6; DB 16; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      30 EAFPDS 35
      |||||
DB      17 EAFPDS 22

Search completed: February 1, 2005, 15:44:50
Job time : 59.778 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:55:01 ; Search time 13.5556 Seconds
(without alignments)
291.016 Million cell updates/sec

Title: SEQ5ASN1967
Perfect score: 41
Sequence: 1 LKGLFTLFAHLVVKPADTL.....QVNISKTDKFAFPDSENDPEK 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 279530

Minimum DB seq length: 21

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :
1: p1r1.*
2: p1r2.*
3: p1r3.*
4: p1r4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	17.1	296	2 A87311	hypothetical prote
2	7	17.1	400	2 C82213	conserved hypotnet
3	7	17.1	422	2 D87550	hypothetical prote
4	7	17.1	453	2 F86846	ABC transporter pe
5	7	17.1	1045	2 B30239	hydroxymethylgluta
6	6	14.6	55	2 E80724	hypothetical prote
7	6	14.6	55	2 B85575	hypothetical prote
8	6	14.6	75	2 F75451	hypothetical prote
9	6	14.6	107	2 B82494	conserved hypotnet
10	6	14.6	108	2 T17826	hypothetical prote
11	6	14.6	120	2 B89932	conserved hypotnet
12	6	14.6	131	2 C83863	hypothetical prote
13	6	14.6	134	2 G69947	phage-related prot
14	6	14.6	162	2 T17044	NADH2 dehydrogenas
15	6	14.6	173	2 S59507	ferric pseudobacti
16	6	14.6	183	2 B75152	adenylyl cyclase r
17	6	14.6	183	2 C71193	hypothetical prote
18	6	14.6	185	2 D95075	conserved hypotnet
19	6	14.6	187	2 C97943	conserved hypotnet
20	6	14.6	190	2 T26019	hypothetical prote
21	6	14.6	191	2 B82903	hypothetical prote
22	6	14.6	201	2 A81380	probable flagellar
23	6	14.6	201	2 B84671	hypothetical prote
24	6	14.6	210	2 B84161	hypothetical prote
25	6	14.6	212	2 A83065	conserved hypotnet
26	6	14.6	212	2 F98221	ycif protein (limp
27	6	14.6	219	2 T38019	urecil phosphoribo
28	6	14.6	220	2 A75287	response regulator
29	6	14.6	220	2 AD0426	two-component syst

30	6	14.6	222	1 B40656	regulatory protein
31	6	14.6	222	1 H65220	transcription regu
32	6	14.6	222	1 AD1022	two-component respo
33	6	14.6	222	2 D86106	hypothetical prote
34	6	14.6	222	2 G91265	hypothetical prote
35	6	14.6	223	2 G95193	ylme protein (limpo
36	6	14.6	223	2 D98060	conserved hypotnet
37	6	14.6	225	2 C97435	transcription acti
38	6	14.6	225	2 AG2653	two component resp
39	6	14.6	226	2 G83293	probable two-compo
40	6	14.6	226	2 T35202	probable two-compo
41	6	14.6	227	1 B47080	transcription acti
42	6	14.6	227	2 F83097	probable two-compo
43	6	14.6	227	2 A90705	probable 2-compone
44	6	14.6	227	2 D85555	probable 2-compone
45	6	14.6	227	2 A64790	probable transcript

ALIGNMENTS

RESULT 1

A87311 hypothetical protein CC0498 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004

C:Accession: A87311

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: A87311

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-296 <STD>

A:Cross-references: UNIPROT:Q9AAU6; GB:AE005673; NID:gt1421677; PIDN:AKK2485.1; GSPDB:GN

C:Genetics:

A:Gene: CC0498

Query Match 17.1%; Score 7; DB 2; Length 296;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 LEFTPAG 10
Db 240 LEFTPAG 246

RESULT 2

C82213 conserved hypothetical protein VC1339 [imported] - Vibrio cholerae (strain N16961 serogrup

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004

C:Accession: C82213

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.

Charidson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoti, I.; Sellers, P

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: C82213

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-400 <HEI>

A:Cross-references: UNIPROT:Q9KSB9; GB:AE004213; GB:AE003852; NID:g9655819; PIDN:AAF94497

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC1339

A:Map position: 1

C:Superfamily: uncharacterized conserved protein

Query Match 17.1%; Score 7; DB 2; Length 400;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 QVNISK 28
Db 139 QVNISK 145

RESULT 3

hypothetical protein CC2429 [imported] - Caulobacter crescentus
C/Species: Caulobacter crescentus
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C/Accession: D87550

R/Merman, W.C.; Feidbylum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Debboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gelin, M.L.; Haft, D.H.; Klot
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A/Title: Complete Genome Sequence of Caulobacter crescentus.
A/Reference number: A87249; MUID:21173698; PMID:11259647

A/Accession: D87550

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-422 <STO>

A/Cross-references: UNIPROT:Q9A5I8; GB:AE005673; NID:G13423972; PIDN:AAK24400.1; GSPDB:C
C/Genetics:

A/Gene: CC2429

Query Match 17.1%; Score 7; DB 2; Length 422;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LEFTFAG 10
Db 204 LEFTFAG 210

RESULT 4

AAC transporter permease protein ysdA [imported] - Lactococcus lactis subsp. lactis (str
C/Species: Lactococcus lactis subsp. lactis
C/Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C/Accession: F66846

R/Bolotin, A.; Winkler, P.; Manger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001

A/Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A/Reference number: A86625; MUID:21235186; PMID:11337471

A/Accession: F66846

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-453 <STO>

A/Cross-references: UNIPROT:Q9CER2; GB:AE005176; PID:G12724797; PIDN:AAK05872.1; GSPDB:C
A/Experimental source: strain IL1403

C/Genetics:

A/Gene: ysdA

Query Match 17.1%; Score 7; DB 2; Length 453;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 QVNISK 28
Db 127 QVNISK 133

RESULT 5

hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) II - Yeast (Saccharomyces cerevisiae)
N/Alternate names: protein L934.2; protein YLR450w
C/Species: Saccharomyces cerevisiae
C/Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 09-Jul-2004
C/Accession: B30239; B24317; S55972

R/Basson, M.E.; Thoresness, M.; Finer-Moore, J.; Stroud, R.M.; Rine, J.
Mol. Cell. Biol. 8, 3797-3808, 1988

A/Title: Structural and functional conservation between yeast and human 3-hydroxy-3-methyl
A/Reference number: A93105; MUID:89127221; PMID:3065625

A/Accession: B30239

A/Molecule type: DNA

A/Residues: 1-1045 <BMS>

A/Cross-references: UNIPROT:P12684; EMBL:M22255; NID:G171687; PIDN:AAA34677.1; PID:G171687

R/Basson, M.E.; Thoresness, M.; Rine, J.
Proc. Natl. Acad. Sci. U.S.A. 83, 5563-5567, 1986

A/Title: Saccharomyces cerevisiae contains two functional genes encoding 3-hydroxy-3-met
A/Reference number: A94112; MUID:86287298; PMID:3526336

A/Accession: B24317

A/Molecule type: DNA

A/Residues: 772-961 <BAS2>

R/Du, Z.
Submitted to the EMBL Data Library, March 1995

A/Description: The sequence of S. cerevisiae cosmid 9324.

A/Reference number: S55966

A/Accession: S55972

A/Molecule type: DNA

A/Residues: 1-1045 <DU2>

A/Cross-references: GB:U22382; NID:G717059; PIDN:AAB67527.1; PID:G717061; MIPS:YLR450w
C/Genetics:

A/Gene: SGD:HM2

A/Cross-references: SGD:S0004442; MIPS:YLR450w

A/Map position: 12R

C/Superfamily: hydroxymethylglutaryl-CoA reductase I
C/Keywords: coenzyme A; oxidoreductase; sterol biosynthesis; transmembrane protein
F/29-45/Domain: transmembrane #status predicted <TM1>

F/248-264/Domain: transmembrane #status predicted <TM2>

F/331-347/Domain: transmembrane #status predicted <TM3>

F/402-418/Domain: transmembrane #status predicted <TM4>

F/502-518/Domain: transmembrane #status predicted <TM5>

F/681-697/Domain: transmembrane #status predicted <TM6>

F/991-1007/Domain: transmembrane #status predicted <TM7>

Query Match 17.1%; Score 7; DB 2; Length 1045;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 HUYKPA 17
Db 10 HUYKPA 16

RESULT 6

hypothetical protein ECG0765 [imported] - Escherichia coli (strain O157:H7, substrain R10

C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C/Accession: E90724

R/Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
Gagaara, N.; Yasunaga, T.; Kohara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno

A/Reference number: A59629; MUID:21156231; PMID:11258796

A/Accession: E90724

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-55 <HAY>

A/Cross-references: UNIPROT:Q8X982; GB:BA000007; PIDN:BA834188.1; PID:G13360224; GSPDB:G

A/Experimental source: strain O157:H7, substrain R10D 0509952

C/Genetics:

A/Gene: ECG0765

Query Match 14.6%; Score 6; DB 2; Length 55;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 TLNQVN 24
Db 36 TLNQVN 41

RESULT 7

E85575
 hypothetical protein Z0896 [imported] - Escherichia coli (strain O157:H7, substrain EDL957)
 C/Species: Escherichia coli
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
 C/Accession: E85575
 R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, B.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A/Reference number: A85480; MUID:21074935; PMID:11206551
 A/Accession: E85575
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-55 <STO>
 A/Cross-references: UNIPROT:Q8X982; GB:AE005174; NID:g12513662; PIDN:AMG55065.1; GSPDB:C
 A/Experimental source: strain O157:H7, substrain EDL957
 C/Genetics:
 A/Gene: Z0896

Query Match 14.6%; Score 6; DB 2; Length 55;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 19 TLNQVN 24
 |||||
 Db 36 TLNQVN 41

RESULT 8

F73451
 hypothetical protein - Deinococcus radiodurans (strain R1)
 C/Species: Deinococcus radiodurans
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
 C/Accession: F73451
 R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 S.; Shen, M.; Vamathevan, J.U.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Me
 Science 286, 1571-1577, 1999
 A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A/Reference number: A75250; MUID:20036896; PMID:10567266
 A/Accession: F73451
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-75 <WHI>
 A/Cross-references: UNIPROT:Q9RVN1; GB:AE001951; GB:AE000513; NID:g6458712; PIDN:AAF1057
 C/Experimental source: strain R1
 C/Genetics:
 A/Gene: DR0996
 A/Map position: 1

Query Match 14.6%; Score 6; DB 2; Length 75;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 7 LFAGHL 12
 |||||
 Db 20 LFAGHL 25

RESULT 9

E82494
 conserved hypothetical protein VCA0152 [imported] - Vibrio cholerae (strain N16961 serog
 C/Species: Vibrio cholerae
 C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
 C/Accession: E82494
 R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
 Chardon, D.; Brimacombe, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoti, I.; Sellers, F
 Nature 406, 477-483, 2000
 A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A/Reference number: A82035; MUID:20406833; PMID:10952301

A/Accession: E82494
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-107 <HEI>
 A/Cross-references: UNIPROT:Q9KN14; GB:AE004356; GB:AE003853; NID:g9657536; PIDN:AAF9606;
 A/Experimental source: serogroup O1, strain N16961, biotype El Tor
 C/Genetics:
 A/Gene: VCA0152
 A/Map position: 2
 C/Superfamily: Pyrococcus horikoshii hypothetical protein PH1428

Query Match 14.6%; Score 6; DB 2; Length 107;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 4 LFTLFA 9
 |||||
 Db 14 LFTLFA 19

RESULT 10

T17826
 hypothetical protein a327R - Chlorella virus PBCV-1
 C/Species: Chlorella virus PBCV-1
 C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C/Accession: T17826
 R/Graves, M.V.; Van Etten, J.L.
 submitted to the EMBL Data Library, May 1999
 A/Reference number: Z18806
 A/Accession: T17826
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-108 <GRA>
 A/Cross-references: UNIPROT:Q84641; EMBL:U42580; NID:g4028896; PIDN:AAC96695.1
 A/Experimental source: specific host Chlorella strain NC64A
 C/Genetics:
 A/Note: a327R
 C/Superfamily: Chlorella virus PBCV-1 hypothetical protein a327R

Query Match 14.6%; Score 6; DB 2; Length 108;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 16 PADTLN 21
 |||||
 Db 36 PADTLN 41

RESULT 11

E89932
 conserved hypothetical protein SA1356 [imported] - Staphylococcus aureus (strain N315)
 C/Species: Staphylococcus aureus
 C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
 C/Accession: E89932
 R/Kurdoz, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi
 ma, A.; Mizutani-Ti, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; I
 Lancet 357, 1225-1240, 2001
 A/Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A/Reference number: A89758; MUID:21311952; PMID:11418146
 A/Accession: E89932
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-120 <KUR>
 A/Cross-references: UNIPROT:Q99TW8; GB:BA000018; PID:g13701324; PIDN:BAH42618.1; GSPDB:GR
 A/Experimental source: strain N315
 C/Genetics:
 A/Gene: SA1356
 C/Superfamily: Bacillus subtilis conserved hypothetical protein yghY
 Query Match 14.6%; Score 6; DB 2; Length 120;
 Best Local Similarity 100.0%; Pred. No. 35;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 VNISKT 28
|||||
DB 81 VNISKT 86

RESULT 12

C83863
hypothetical protein BH1707 [imported] - *Bacillus halodurans* (strain C-125)
C/Species: *Bacillus halodurans*
C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C/Accession: C83863
R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A/Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
A/Reference number: A83650; MUID:20512582; PMID:11058132
A/Accession: C83863
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-131 <STO>
A/Cross-references: UNIPROT:O9KC67; GB:AP001512; GB:BA000004; NID:g10174030; PIDN:BA054
A/Experimental source: strain C-125
C/Genetics:
A/Gene: BH1707

Query Match 14.6%; Score 6; DB 2; Length 131;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VKFPAD 18
|||||
DB 45 VKFPAD 50

RESULT 13

G69947
phage-related protein homolog yqbl - *Bacillus subtilis*
C/Species: *Bacillus subtilis*
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C/Accession: G69947
R/Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Aevedo, V.; Berter
C.; Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrati, E.
Nature 390, 249-256, 1997
A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
lech, J.; Harwood, C.R.; Henuit, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidos, A.; Lardinois
A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y.M.; Ogawa, K.; Ogilvie, A.; Oudega, B.; Park, S.H.; Parro, V.; Sadale, Y.; Sato, T.; Scallan
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scallan
A/Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot
akenuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A/Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A/Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A/Reference number: A69580; MUID:98044033; PMID:9384377
A/Accession: G69947
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-134 <KUN>
A/Cross-references: UNIPROT:P45928; GB:Z99117; GB:AL009126; NID:92634966; PIDN:CAB14548
A/Experimental source: strain 168
C/Genetics:
A/Gene: yqbl

Query Match 14.6%; Score 6; DB 2; Length 134;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 FDSEND 38
|||||
DB 95 FDSEND 100

RESULT 14

T17044
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - common snowdrop chloroplast (frag
C/Species: *Chloroplast Galanthus nivalis* (common snowdrop)
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C/Accession: T17044
R/Graham, S.W.; Barrett, S.C.H.
submitted to the EMBL Data Library, November 1996
A/Description: Phylogenetic analysis of Narcissus L. (Amaryllidaceae) based on the chlor
A/Reference number: Z17939
A/Accession: T17044
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: mRNA
A/Residues: 1-162 <GRA>
A/Cross-references: UNIPROT:O47204; EMBL:U79222; NID:92827789; PID:92827790; PIDN:AAB998
C/Genetics:
A/Genome: chloroplast
A/Note: nhf
C/Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C/Keywords: chloroplast; membrane-associated complex; NAD; oxidoreductase

Query Match 14.6%; Score 6; DB 2; Length 162;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GLFTLF 8
|||||
DB 63 GLFTLF 68

RESULT 15

S59507
ferric pseudobactins receptor protein RFS - *Pseudomonas putida* (fragment)
C/Species: *Pseudomonas putida*
C/Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 13-Mar-1997
C/Accession: S59507
R/Koster, M.; Ova, W.; Bitter, W.; Weisbeek, P.
Mol. Gen. Genet. 248, 735-743, 1995
A/Title: Multiple outer membrane receptors for uptake of ferric pseudobactins in *Pseudom*
A/Reference number: S59503; MUID:96069713; PMID:7476877
A/Accession: S59507
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-173 <KOS>

Query Match 14.6%; Score 6; DB 2; Length 173;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GLFTLF 8
|||||
DB 98 GLFTLF 103

Search completed: February 1, 2005, 15:33:04
Job time: 13.5556 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:54:06 ; Search time 70.5556 Seconds
(without alignments)
334.352 Million cell updates/sec

Title: SEQ5ALA1854
Perfect score: 41
Sequence: 1 EKWKKNHMGPFMSILOEHIG.....MKKEELTSHQSQTAFLEA 41

Scoring table: OLIGO
Gapop 60.0 , Gapect 60.0

Searched: 1825181 seqs, 575374646 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1810864

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : UniProt 02.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	349	2	Q96ES5
2	41	100.0	897	2	Q8N7L7
3	41	100.0	1106	2	Q6P197
4	41	100.0	1106	2	AAH65205
5	40	48.8	958	1	BP28_MACFA
6	20	48.8	2144	1	BP28_HUMAN
7	19	46.3	349	2	Q8VCK1
8	19	46.3	408	2	Q8CCT5
9	19	46.3	163	2	Q8B1C5
10	19	46.3	354	2	Q8B1C2
11	19	46.3	445	2	Q7N5B2
12	19	46.3	448	2	Q7N5B2
13	19	46.3	453	2	Q7N5B2
14	19	46.3	454	2	Q8B1C2
15	19	46.3	454	2	Q8B1C2
16	19	46.3	454	2	Q8B1C2
17	19	46.3	454	2	Q8B1C2
18	19	46.3	454	2	Q8B1C2
19	19	46.3	454	2	Q8B1C2
20	19	46.3	454	2	Q8B1C2
21	19	46.3	454	2	Q8B1C2
22	19	46.3	454	2	Q8B1C2
23	19	46.3	454	2	Q8B1C2
24	19	46.3	454	2	Q8B1C2
25	19	46.3	454	2	Q8B1C2
26	19	46.3	454	2	Q8B1C2
27	19	46.3	454	2	Q8B1C2
28	19	46.3	454	2	Q8B1C2
29	19	46.3	454	2	Q8B1C2
30	19	46.3	454	2	Q8B1C2
31	19	46.3	454	2	Q8B1C2

32	8	19.5	488	2	Q07440	007440 pseudomonas
33	8	19.5	488	2	Q7DC90	Q7DC90 pseudomonas
34	8	19.5	493	2	Q92567	Q92567 rhizobium m
35	7	17.1	258	2	Q88E22	Q88E22 pseudomonas
36	7	17.1	417	2	Q9NVT4	Q9NVT4 homo sapien
37	7	17.1	460	2	Q723N1	Q723N1 trichoderma
38	7	17.1	464	2	Q8NKT5	Q8NKT5 acidianus a
39	7	17.1	482	2	Q8XQ15	Q8XQ15 ralsiconia s
40	7	17.1	525	2	Q8GZT5	Q8GZT5 emiliania h
41	7	17.1	525	2	Q88UT6	Q88UT6 lactobacill
42	7	17.1	537	1	CYDA_AZOVI	Q09049 azotobacter
43	7	17.1	545	2	Q6B055	Q6B055 debaryomyce
44	7	17.1	667	2	Q9KX28	Q9KX28 streptococc
45	7	17.1	739	2	Q9RQT4	Q9RQT4 streptococc

ALIGNMENTS

RESULT 1

Q96ES5 PRELIMINARY; PRT; 349 AA.

AC Q96ES5; 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE FLJ10359 protein.
GN Name=FLJ10359;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=23388257; PubMed=12477932;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Spelman C.M., Schuler G.D.,
RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang J., Hsieh F.,
RA Datchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Baha S.S., Loguelfano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.U., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalke D.E., Schermer A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RA Strauberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC011983; AAL11983.1; -
DR InterPro: IPR008938; ARM.
SQ SEQUENCE 349 AA; 39921 MW; 3A359597FF7079EB CRC64;

Query Match 100.0%; Score 41; DB 2; Length 349;
Best Local Similarity 100.0%; Pred. No. 1.9e-36;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKWKKNHMGPFMSILOEHIGAMKKEELTSHQSQTAFLEA 41
DB 39 EKWKKNHMGPFMSILOEHIGAMKKEELTSHQSQTAFLEA 79

```

RESULT 2
08N7L7 PRELIMINARY; PRT; 897 AA.
ID 08N7L7;
AC 08N7L7;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein FLJ40893.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RX PubMed=14702039;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa B., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Niimiyama K., Itoh Y., Yamashita H., Murakawa M., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hirooka S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yoshida M., Houchu T., Kusano Y.,
RA Kanehori K., Takahashi-Fujii A., Hara R., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotaka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki M., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake K., Inagaki H., Ikemura Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs";
RL Nat. Genet. 36:40-45 (2004).
DR EMBL; AK098212; BAC05261.1;
DR InterPro; IPR008938; ARM.
SQ SEQUENCE 897 AA; 101574 MW; 5FF6A94FE8855895 CRC64;

Query Match 100.0%; Score 41; DB 2; Length 897;
Best Local Similarity 100.0%; Pred. No. 4,4e-36;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 EKNWKNHMGPFMSILQEHIGAMKKEELTSHOSQUTAFPLEA 41
Db 587 EKNWKNHMGPFMSILQEHIGAMKKEELTSHOSQUTAFPLEA 627

RESULT 3
06P197 PRELIMINARY; PRT; 1106 AA.
ID 06P197;
AC 06P197;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE FLJ10359 protein (Fragment).
GN Name=FLJ10359;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

```

NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Tothiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA Strausberg R.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC065205; AAH65205.1;
DR InterPro; IPR008938; ARM.
FT NON TER 1
SQ SEQUENCE 1106 AA; 125359 MW; 09F7CE94042302C4 CRC64;

Query Match 100.0%; Score 41; DB 2; Length 1106;
Best Local Similarity 100.0%; Pred. No. 5.3e-36;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 EKNWKNHMGPFMSILQEHIGAMKKEELTSHOSQUTAFPLEA 41
Db 796 EKNWKNHMGPFMSILQEHIGAMKKEELTSHOSQUTAFPLEA 836

RESULT 4
AAH65205 PRELIMINARY; PRT; 1106 AA.
ID AAH65205;
AC AAH65205;
DT 02-MAR-2004 (TREMBlrel. 27, Created)
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
DE 02-MAR-2004 (TREMBlrel. 27, Last annotation update)
DE FLJ10359 protein (Fragment).
GN Name=FLJ10359;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Tothiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA Strausberg R.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC065205; AAH65205.1;
DR InterPro; IPR008938; ARM.
FT NON TER 1
SQ SEQUENCE 1106 AA; 125359 MW; 09F7CE94042302C4 CRC64;

```


DR EMBL; AK001221; BAA91564.1; ALT_INIT.
 DR SWISS-2DPAGE; Q9H583; HUMAN.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR000357; HEAT.
 DR PROSITE; PS50077; HEAT_REPEAT; FALSE_NEG.
 KM Polymorphism.
 FT REPEAT 2106 HEAT.
 FT VARIANT 1694 N -> S.
 FT VARIANT 1694 /FTID=VAR_010939.
 FT VARIANT 1854 V -> A.
 FT VARIANT 1967 /FTID=VAR_010940.
 FT VARIANT 1967 N -> D.
 FT VARIANT 2017 /FTID=VAR_010941.
 FT VARIANT 2017 E -> G.
 FT VARIANT 2017 /FTID=VAR_010942.
 SQ SEQUENCE 2144 AA; 242355 MW; D66816EE78D8C9B7 CRC64;
 Query Match 46.8%; Score 20; DB 1; Length 2144;
 Best Local Similarity 100.0%; Pred. No. 1e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 EKNWKNHMGPPMSILQEHIG 20
 Db 1834 EKNWKNHMGPPMSILQEHIG 1853
 RESULT 7
 Q8VCK1 PRELIMINARY; PRT; 349 AA.
 AC Q8VCK1
 DT 01-MAR-2002 (TRENBLrel. 20. Created)
 DT 01-MAR-2002 (TRENBLrel. 20. Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26. Last annotation update)
 DE BC019693 protein.
 GN Name=BC019693;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=mx FVB/N;
 RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Straube R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stepieton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshlyuk S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
 RA Krzywinski M.I., Skalska U., Small D.E., Scherch A., Schein J.E.,
 RA Jones S.D., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=mx FVB/N;
 RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
 RA Straube R.L.,
 RT Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC019693; AAH19693.1;
 DR MGI; MGI:2384983; BC019693.
 DR InterPro; IPR008938; ARM.

SQ SEQUENCE 349 AA; 40166 MW; 9763D031AD0F515 CRC64;
 Query Match 46.3%; Score 19; DB 2; Length 349;
 Best Local Similarity 100.0%; Pred. No. 2.5e-12;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 KNWKNHMGPPMSILQEHIG 20
 Db 40 KNWKNHMGPPMSILQEHIG 58
 RESULT 8
 Q8CCT5 PRELIMINARY; PRT; 408 AA.
 AC Q8CCT5
 DT 01-MAR-2003 (TRENBLrel. 23. Created)
 DT 01-MAR-2003 (TRENBLrel. 23. Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26. Last annotation update)
 DE Mus musculus adult male olfactory brain cDNA, RIKEN full-length
 DE enriched library, clone:643040D06 product:hypothetical ARM repeat.
 DE structure containing protein, full insert sequence.
 GN Name=BC019693;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
 RX MEDLINE=9279253; PubMed=10349636;
 RA Carninci P., Hayashizaki Y.,
 RT "High-efficiency full-length cDNA cloning."
 RT Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
 RX MEDLINE=21085660; PubMed=11217851;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RT Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
 RA The FANTOM Consortium;
 RT "The RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RT Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
 RX MEDLINE=20499374; PubMed=11042159;
 RA Carninci P., Shibata Y., Hayatsu N., Suganara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
 RT "Normalization and subcloning of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes."
 RT Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
 RX MEDLINE=20530913; PubMed=11076861;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsuno H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Onaka E., Wataniki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsunura S., Kawai U.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer."
 RT Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.

```

RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hirooka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishii K., Nomura K., Numazaki R., Ohno M., Ohnato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Saeki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AK032134; BAC27721.1; -.
DR MGD; MGI:2384983; BC019693.
DR InterPro; IPR008938; ARM.
KW Hypothetical protein.
SQ
SEQUENCE 408 AA; 46688 MW; 42AB9B13CA3FE67 CRC64;

Query Match 46.3%; Score 19; DB 2; Length 408;
Best Local Similarity 100.0%; Pred. No. 2,9e-12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KMWKMGPFMSIQEHIG 20
Db 99 KMWKMGPFMSIQEHIG 117

RESULT 9
08B1J2 PRELIMINARY; PRT; 163 AA.
ID 08B1J2
AC 08B1J2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 25, Last sequence update)
DE Mus musculus B6-derived CD11+ve dendritic cells cDNA, RIKEN full-
DE length enriched library, clone: F730048A22 product: weakly similar to
DE BA93316.3 (NOVEL KRAB BOX CONTAINING ZINC FINGER GENE).
GN Name=BB114266;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RC MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RC MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA The FANTOM Consortium;
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RC MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";

```

```

RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RC MEDLINE=20530913; PubMed=11076861;
RX Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Kono H., Akiyama Y., Nishi K., Kitanishi T., Tashiro H., Itoh A.,
RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kaishiki M.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hirooka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishii K., Nomura K., Numazaki R., Ohno M., Ohnato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Saeki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
RL Submitted (APR-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AK089532; BAC40916.1; -.
DR HSSP; P15822; IBBO.
DR MGD; MGI:2143362; BB114266.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf_C2H2; 5.
DR ProDom; PD000003; Znf_C2H2; 5.
DR SMART; SM00355; Znf_C2H2; 5.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 5.
SQ SEQUENCE 163 AA; 19461 MW; A1A8FD2BCD74C044 CRC64;

Query Match 19.5%; Score 8; DB 2; Length 163;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 29 SHOSQLTN 36
Db 116 SHOSQLTN 123

RESULT 10
08B1J2 PRELIMINARY; PRT; 354 AA.
ID 08B1J2
AC 08B1J2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mus musculus ES cells cDNA, RIKEN full-length enriched library,
DE clone: C310026E23 product: weakly similar to BA93316.3 (NOVEL KRAB BOX
DE CONTAINING ZINC FINGER GENE).
GN Name=BB114266;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RC MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;

```

RA "High-efficiency full-length cDNA cloning."
 RL Mech. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=21085660; PubMed=11217851;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RA The FANTOM Consortium,
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RL 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=20499374; PubMed=11042159;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RT Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
 RL "Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes."
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=20530913; PubMed=11076861;
 RA Shibata K., Itoh M., Mizawa K., Nagao S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama J., Nishi K., Kitanai T., Taahiro H., Itoh M.,
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Masunoto H., Sakaguchi S., Ikegami T., Kashinagi K.,
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Wataniki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RL sequencing pipeline with 384 multicapillary sequencer."
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RA Aichi J., Mizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Putano M., Hanganaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hirooka T., Hirozane T.,
 RA Hori F., Imocani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
 RA Kurihara C., Matsumura T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numasaki R., Ono M., Ohnato N., Okazaki Y.,
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tomaru A., Takahashi F., Takahashi A., Muramatsu M., Hayashizaki Y.,
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK049344; BAC3636.1; -
 DR HSSP: P08046; IP21.
 DR MGD: MGI:2143362; BHL14266.
 DR GO: GO:0005634; C:nucleus; IEA.
 DR GO: GO:0003676; F:nucleic acid binding; IEA.
 DR GO: GO:0008072; F:zinc ion binding; IEA.
 DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro: IPR001909; KRAE.
 DR InterPro: IPR007087; Znf_C2H2.
 DR Pfam: PF01352; KRAE; 1.
 DR Pfam: PF00096; ZF-C2H2; 7.
 DR ProDom: PDD00003; Znf_C2H2; 6.
 DR SMART: SM00349; KRAE; 1.
 DR SMART: SM00355; ZNF_C2H2; 7.
 DR PROSITE: PS00805; KRAE; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 7.
 DR SEQUENCE 354 AA; 41447 MW; F2C3DDEA21B43B4 CRC64;

```

Query Match      19.5%; Score 8; DB 2; Length 354;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      29 SHOSQOLTA 36
      |||||
      307 SHOSQOLTA 314

Db

RESULT 11
ID Q7NSB2      PRELIMINARY; PRT; 445 AA.
AC Q7NSB2;
DT 01-MAR-2004 (TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Similar to quinol oxidase subunit 1.
OS OrderedLocusNames=plu2043;
OC Photobacterium luntense (subsp. laumondii).
CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC Enterobacteriaceae; Photobacterius.
CX NCBI_TaxID=141679;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TT01;
RX MEDLINE=22957627; PubMed=14528314;
RA Duchand E., Rusnik C., Frangoul L., Buchrieser C., Givaudan A.,
RA Teourit S., Bocs S., Bournaud-Eude C., Chandler M., Charles J.-F.,
RA Dassa E., Derose R., Derzelle S., Freysinet G., Gaudriault S.,
RA Meignie C., Landis A., Powell K., Stiguer F., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;
RT "The genome sequence of the entomopathogenic bacterium Photobacterius
luntense";
RL Nat. Biotechnol. 21:1307-1313(2003).
DR EMBL; BX571865; CAE14336.1; -.
DR Photolyst; plu2043; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002585; Bac_Ibdq_Cox.
DR Pfam; PF01654; Bac_Ibdq_Cox; 1.
KW Complete proteome.
SQ SEQUENCE 445 AA; 49978 MW; BD5F82F209B973FA CRC64;

Query Match      19.5%; Score 8; DB 2; Length 445;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      34 LTAFPLEA 41
      |||||
      103 LTAFPLEA 110

Db

RESULT 12
ID Q7NRX0      PRELIMINARY; PRT; 448 AA.
AC Q7NRX0;
DT 01-MAR-2004 (TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Cyanide insensitive terminal oxidase (EC 1.10.3.-).
OS Name=cioA; OrderedLocusNames=CV3658;
OC Chromobacterium violaceum.
CC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
CC Neisseriaceae; Chromobacterium.
CX NCBI_TaxID=536;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14272 / DSM 30191;
RX MEDLINE=22882880; PubMed=14500782;
RA Vasconcelos A.T.R., de Almeida D.F., Hungria M., Guimaraes C.T.,
RA Antonio R.V., Almeida F.C., de Almeida L.G.P., de Almeida R.,

```

RA Alves-Gomes J.A., Andrade E.M., Arraixe J., de Araujo M.F.F.,
 RA Azeiteiro-Filho S., Azevedo V., Baptista A.J., Batista L.A.M.,
 RA Bactelli J.S., Belo A., van den Berg C., Bogo M., Bonetto S.,
 RA Bordignon J., Brigido M.M., Brito C.A., Brocchi M., Burtly H.A.,
 RA Carvalho A.A., Cardoso D.D.P., Carneiro N.P., Carraro D.M.,
 RA Crezynski-Pasa T.B., Cunha-Junior N.C., Fagundes L.M.O.,
 RA Pantiuchi F., Farías I.P., Felipe M.S.S., Ferrari L.P., Ferro J.A.,
 RA Patro M.T., Franco G.R., Freitas N.S.A., Furlan L.R.,
 RA Garzinielli R.T., Gomes E.A., Gonçalves P.R., Grangeiro T.B.,
 RA Gracetiagalia D., Grissard E.C., Hanna E.S., Jardim S.N., Laurino J.,
 RA Leol L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.P.,
 RA Medeira H.M.F., Mantio G.P., Maranhão A.Q., Martins W.S.,
 RA di Mauro S.M.Z., de Medeiros S.R.B., Meisner R.V., Moreira M.A.M.,
 RA Macielmento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,
 RA Patxao R.P.C., Parente J.A., Pedrosa F.O., Pena S.D.O., Pereira J.O.,
 RA Pereira M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Potrich D.P.,
 RA Ramalho-Neto C.E., Reis A.M.M., Rigó L.V., Rondinelli E.,
 RA Santos E.B.P., Santos F.R., Schneider M.P.C., Senauez H.N.,
 RA Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,
 RA Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,
 RA Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Umenya T.,
 RA Valtore A., Wasserm R., Zana A., Simpson A.J.G.,
 RA "The complete genome sequence of Chromobacterium violaceum reveals
 RT remarkable and exploitable bacterial adaptability.",
 RL Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).
 DR EMBL: AB016923; AAC01320.1; -
 DR GO: GO:0016020; C:membrane; IEA.
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.
 DR GO: GO:0006118; P:electron transport; IEA.
 DR InterPro: IPR002585; Bac_Ubq_Cox.
 DR Pfam: PF01654; Bac_Ubq_Cox; I.
 KM Complete proteome; Oxidoreductase.
 SQ SEQUENCE 448 AA; 49586 MW; 0803368EB45024D0 CRC64;

Query Match 19.5%; Score 8; DB 2; Length 448;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 LTAFFLEA 41
 Db 104 LTAFFLEA 111

RESULT 13
 Q98DW1 PRELIMINARY; PRT; 453 AA.
 AC Q98DW1;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Quinol oxidase subunit I.
 GN OrderedLocustNames=mlr4524;
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 CC Phyllobacteriaceae; Mesorhizobium.
 CX NCBI_TaxID=381;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082936; PubMed=11214974;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti (supplement).",
 RT DNA Res. 7:381-406(2000).
 RL [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;

RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti.",
 RT DNA Res. 7:331-338(2000).
 DR EMBL: AP003004; BAB51159.1; -
 DR GO: GO:0016020; C:membrane; IEA.
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.
 DR GO: GO:0006118; P:electron transport; IEA.
 DR InterPro: IPR002585; Bac_Ubq_Cox.
 DR Pfam: PF01654; Bac_Ubq_Cox; I.
 KM Complete proteome.
 SQ SEQUENCE 453 AA; 49427 MW; 1D2439B6E34BFA0F CRC64;

Query Match 19.5%; Score 8; DB 2; Length 453;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 LTAFFLEA 41
 Db 100 LTAFFLEA 107

RESULT 14
 Q87H28 PRELIMINARY; PRT; 454 AA.
 AC Q87H28;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Cytochrome BD2, subunit I.
 GN OrderedLocustNames=VPA1137;
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 CC Vibrionaceae; Vibrio.
 CX NCBI_TaxID=670;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=RIMD 221063 / serotype O3:K6;
 RX MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Nejima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.,
 RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
 RT distinct from that of V. cholerae.",
 RL Lancet 361:743-749(2003).
 DR EMBL: AP005088; BAC62480.1; -
 DR GO: GO:0016020; C:membrane; IEA.
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.
 DR GO: GO:0006118; P:electron transport; IEA.
 DR InterPro: IPR002585; Bac_Ubq_Cox.
 DR Pfam: PF01654; Bac_Ubq_Cox; I.
 KM Complete proteome.
 SQ SEQUENCE 454 AA; 50975 MW; 318223526BBAF46 CRC64;

Query Match 19.5%; Score 8; DB 2; Length 454;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 LTAFFLEA 41
 Db 103 LTAFFLEA 110

RESULT 15
 Q7WB10 PRELIMINARY; PRT; 462 AA.
 AC Q7WB10;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

```

PT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cytochrome oxidase subunit I.
GN Name=ci0A; Synonyms=gx1A; OrderedLocusNames=BPP1024;
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP
RC STRAIN=12822 / ATCC BAA-587;
RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
RA Parkhill J., Sebatina M., Preston A., Murphy L.D., Thomson N.R.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble S., Hamlin N., Hauser H., Holtroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL: BX640426; CAE36324.1; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0006118; P:electron transport; IEA.
DR InterPro: IPR002585; Bac_Ubq_Cox.
DR Pfam: PF01654; Bac_Ubq_Cox; 1.
KW Complete proteome.
SQ
SEQUENCE 462 AA; 49902 MW; 631FAC87419163ED CRC64;

Query Match 19.5%; Score 8; DB 2; Length 462;
Best Local Similarity 100.0%; Pred.No. 3.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 34 LTAPFLEA 41
   |||||
Db 103 LTAPFLEA 110

```

Search completed: February 1, 2005, 15:31:01
 Job time : 72.5556 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:54:06 ; Search time 70.5556 Seconds

(without alignments)
334.352 Million cell updates/sec

Title: SEQ5_2030_2070

Sequence: 1 RLGGSEKFORVTKHLIPCI.....QFSVAMADSLMKPLNYQIL 41

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1825181 seqs, 575374646 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1810864

Minimum DB seq length: 21

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

UniProt_02:.*
1: uniprot_sprot:.*
2: uniprot_trembl:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	349	2	Q96ES5
2	41	100.0	349	2	Q96ES5
3	41	100.0	1106	2	Q8N7L7
4	41	100.0	1106	2	Q6P197
5	41	100.0	1106	2	AAH65205
6	26	63.4	958	1	BP28_HUMAN
7	14	34.1	349	2	Q8VCK1
8	14	34.1	408	2	Q8CCT5
9	7	17.1	97	2	Q7XPO6
10	7	17.1	285	2	Q9N045
11	7	17.1	365	2	Q8OV55
12	7	17.1	384	2	Q7TMO4
13	7	17.1	625	2	Q6IRCS
14	7	17.1	625	2	AAH70973
15	7	17.1	639	2	Q29993
16	7	17.1	741	2	Q6NZG9
17	7	17.1	741	2	AAH66140
18	7	17.1	1076	2	Q7VMH1
19	7	17.1	1076	2	Q7B61
20	6	14.6	88	2	Q6TV79
21	6	14.6	88	2	Q6Z8M3
22	6	14.6	88	2	BAH10078
23	6	14.6	90	2	Q9X5L8
24	6	14.6	96	2	Q828R1
25	6	14.6	97	2	P70723
26	6	14.6	102	2	Q8WY95
27	6	14.6	106	2	Q8LFE2
28	6	14.6	106	2	Q9SYE2
29	6	14.6	108	2	Q6IVM8
30	6	14.6	108	2	AAH38620
31	6	14.6	113	1	RL24_MICU

32	6	14.6	133	2	Q6LL94	Q6L194 photobacter
33	6	14.6	133	2	CAG22120	Cag22120 photobact
34	6	14.6	135	1	E314_ADE07	P5135 human adeno
35	6	14.6	135	2	Q775K7	Q775K7 human adeno
36	6	14.6	135	2	Q775L5	Q775L5 human adeno
37	6	14.6	135	2	Q9PX47	Q9PX47 human adeno
38	6	14.6	135	2	Q91CK8	Q91CK8 human adeno
39	6	14.6	135	2	AAH89974	AAH89974 human ade
40	6	14.6	136	2	AAH89974	AAH89974 human ade
41	6	14.6	143	2	Q6E779	Q6E779 saprolegnia
42	6	14.6	154	2	Q8NC54	Q8NC54 homo sapien
43	6	14.6	161	2	Q8C908	Q8C908 mus musculu
44	6	14.6	161	2	Q8CAT8	Q8CAT8 mus musculu
45	6	14.6	163	2	Q710P2	Q710P2 oryza sativ

ALIGNMENTS

RESULT 1

ID	Q96ES5	PRELIMINARY	PRT	349 AA.
AC	Q96ES5			
DT	01-DEC-2001 (TREMBLrel. 19, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)			
DE	FLJ10359 protein.			
CN	Name=FLJ10359;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
NC	NCBI_Taxid=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Ovary;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,			
RA	Brownstein M.J., Ueda T.B., Toshitsugu S., Carrinchi P., Prange C.,			
RA	Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,			
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.U., Hulik S.W.,			
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Rodriguez A.C., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Blakesley R.M., Schmutz J., Myer R.M., Buterfield Y.S.,			
RA	Krzywinski M.I., Skalska U., Smalins D.E., Schermer A., Schein J.E.,			
RT	"Generation and initial analysis of more than 15,000 full-length human			
RT	and mouse cDNA sequences."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Ovary;			
RA	Strauberg R.L.,			
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC011983; AAH11983.1; -			
DR	InterPro; IPR006938; ARM.			
SO	SEQUENCE 349 AA; 39921 MW; 3A359597FF70799B CRC64;			

Query Match 100.0%; Score 41; DB 2; Length 349;
Best Local Similarity 100.0%; Pred. No. 1.6e-36;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	RLGGSEKFORVTKHLIPCIAGSVAMADSLMKPLNYQIL 41
DB	235	RLGGSEKFORVTKHLIPCIAGSVAMADSLMKPLNYQIL 275

```

RESULT 2
ID 08N7L7 PRELIMINARY; PRT; 897 AA.
AC 08N7L7;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein FLJ40893.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RX PubMed=14702039;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Makamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Ohtsuka M., Nishitani T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto K., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Niinomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Ono Y., Takiguchi S., Watanabe M., Hirooka S., Chiba Y., Ishida S.,
RA Kanohori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotaka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Taniguchi A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujimori T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo K., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake K., Inagaki H., Ikemura Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RL Nat. Genet. 36:40-45(2004).
DR EMBL; AK098212; BAC05261.1; -
DR InterPro; IPR008938; ARM.
SQ SEQUENCE 897 AA; 101574 MW; 5FF6A94FE8855895 CRC64;

Query Match 100.0%; Score 41; DB 2; Length 897;
Best Local Similarity 100.0%; Pred. No. 3.5e-36;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 RLGGSEKQERVTYKHLIPICIAQPSVAMADDSLMKPIYQIL 41
Db 783 RLGGSEKQERVTYKHLIPICIAQPSVAMADDSLMKPIYQIL 823

RESULT 3
ID 06P197 PRELIMINARY; PRT; 1106 AA.
AC 06P197;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE FLJ10359 protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

```

```

NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Shat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marulita K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshitsuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Villalón D.K., Wozny K.C., Hale S., Garcia A.M., Gay L.V., Hulyk S.W.,
RA Villalón D.K., Wozny K.C., Hale S., Garcia A.M., Gay L.V., Hulyk S.W.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Kravynski M.I., Skalska U., Smalins D.E., Schmech A., Schein J.E.,
RA Jones S.J., Maira M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN 12
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA Strausberg R.;
DR EMBL; BC065205; AAH65205.1; -
DR InterPro; IPR008938; ARM.
SQ SEQUENCE 1106 AA; 125359 MW; 09F7CE94042302C4 CRC64;

Query Match 100.0%; Score 41; DB 2; Length 1106;
Best Local Similarity 100.0%; Pred. No. 4.1e-36;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 RLGGSEKQERVTYKHLIPICIAQPSVAMADDSLMKPIYQIL 41
Db 992 RLGGSEKQERVTYKHLIPICIAQPSVAMADDSLMKPIYQIL 1032

RESULT 4
ID AAH65205 PRELIMINARY; PRT; 1106 AA.
AC AAH65205;
DT 02-MAR-2004 (TREMBlrel. 27, Created)
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
DE 02-MAR-2004 (TREMBlrel. 27, Last annotation update)
DE FLJ10359 protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Shat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marulita K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshitsuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Villalón D.K., Wozny K.C., Hale S., Garcia A.M., Gay L.V., Hulyk S.W.,
RA Villalón D.K., Wozny K.C., Hale S., Garcia A.M., Gay L.V., Hulyk S.W.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

```

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimmon J., Schmitt J., Myers R.M., Butterfield Y.S.,
 RA Krzywninski M.I., Skalska U., Smallos D.E., Scherch A., Schain J.E.,
 RA Jones S.U., Mair M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph.
 RA Strauberg R.,
 RL Submitted (JAN-2004) to the EMBL/Genbank/DBJ databases.
 DR EMBL; BC065205; AAH65205.1; -
 FT NON TER 1
 SQ SEQUENCE 1106 AA; 125359 MW; 09F7CE94042302C4 CRC64;
 Query Match 100.0%; Score 41; DB 2; Length 1106;
 Best Local Similarity 100.0%; Pred. No. 4.1e-36;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 RLGGEEKQERVTGHLIPCIAPSVAMADSLMKPLNYQIL 41
 Db 992 RLGGEEKQERVTGHLIPCIAPSVAMADSLMKPLNYQIL 1032
 RESULT 5
 BP28 HUMAN STANDARD; PRT; 2144 AA.
 ID Q9H583; Q9NM23;
 AC Q9H583; Q9NM23;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Protein BAP28.
 GN Name=BAP28;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. AND VARIANTS SER-1694; ALA-1854; ASP-1967 AND
 RP GLY-2017.
 RA Bouguieret L., Chumakov I., Barry C., Cohen-Akennine A.;
 RT "A novel BAP28 gene and protein."
 RL Patent number WO0100669, 04-JAN-2001.
 RN [2]
 RP SEQUENCE OF 1534-2144 FROM N.A.
 RA Cobley V.;
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1777-2144 FROM N.A.
 RX PubMed14702039; DOI=10.1038/ng1285;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Wakematsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
 RA Nagahata K., Murakami K., Yasuda T., Iwayanagi T., Nagatsuma M.,
 RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
 RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
 RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa K.,
 RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
 RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hirao K., Chiba Y.,
 RA Iehida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Houchi T.,
 RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
 RA Nomura Y., Togliya S., Komai F., Hara R., Takeuchi K., Arita M.,
 RA Imose N., Mueseshino K., Yuki H., Oshima A., Sasaki N., Aotsuka S.,
 RA Yoshihaka Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
 RA Moriya S., Moriwaya H., Satoh N., Takami S., Taraibima Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 RA Hibiigaki H., Watanabe T., Sugiyama A., Takamoko S., Fukuzumi Y.,
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujizawa T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,

RA Kawabata A., Hiki T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Ohtsuna K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
 RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Maehuo Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
 RT "Complete sequencing and characterization of 21,243 full-length human
 RT cDNAs.";
 RL Nat. Genet. 36:40-45(2004).
 CC -1- SIMILARITY: Belongs to the BAP28 family.
 CC -1- SIMILARITY: Contains 1 HEAT repeat.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ldb-rib.ch/announce/>
 CC or send an email to license@ldb-rib.ch).
 CC -----
 DR EMBL; AX067150; CAC26776.1; -
 DR EMBL; AL136105; CAC15948.1; -
 DR EMBL; AK001221; BAA91564.1; ALT_INIT.
 DR SWISS-2DPAGE; Q9H583; HUMAN.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR000357; HEAT.
 DR PROSITE; PS50077; HEAT_REPEAT; FALSE_NEG.
 DR Polymorphism.
 FT REPEAT 2106 2142 HEAT.
 FT VARIANT 1694 1694 N->S.
 FT VARIANT 1854 1854 /FTID=VAR_010939.
 FT VARIANT 1967 1967 V->A.
 FT VARIANT 1967 1967 /FTID=VAR_010940.
 FT VARIANT 2017 2017 N->D.
 FT VARIANT 2017 2017 /FTID=VAR_010941.
 FT VARIANT 2017 2017 E->G.
 FT SEQUENCE 2144 AA; 242355 MW; D66816E78D8C8B7 CRC64;
 Query Match 100.0%; Score 41; DB 1; Length 2144;
 Best Local Similarity 100.0%; Pred. No. 7.1e-36;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 RLGGEEKQERVTGHLIPCIAPSVAMADSLMKPLNYQIL 41
 Db 2030 RLGGEEKQERVTGHLIPCIAPSVAMADSLMKPLNYQIL 2070
 RESULT 6
 BP28 MACFA STANDARD; PRT; 958 AA.
 ID Q9GWA4;
 AC Q9GWA4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Protein BAP28 (QnpA-17571) (Fragment).
 GN Name=BAP28;
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Osada N., Hida M., Kusuda J., Tanuma R., Ieeki K., Hirai M., Terao K.,
 RA Suzuki Y., Sugano S., Hashimoto K.;
 RT "Isolation of full-length cDNA clones from macaque brain cDNA
 RT libraries.";
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: Belongs to the BAP28 family.
 CC -1- SIMILARITY: Contains 1 HEAT repeat.

```
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB049842; BAB16728.1; ALT_INT.
DR InterPro: IPR008938; ARM.
DR InterPro: IPR000357; HEAT.
DR PROSITE: PSS0077; HEAT_REPEAT; FALSE_NEG.
FT NON TER 1 956
FT REPEAT 920 956
SQ SEQUENCE 958 AA; 108644 MW; 3DBD95C3623C6B31 CRC64;

Query Match 63.4%; Score 26; DB 1; Length 958;
Best Local Similarity 100.0%; Pred. No. 1.1e-19;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 PCIAQSFVAMADDSLMKPLNYQIL 41
DB 859 LIPCIQSFVAMADDSLMKPLNYQIL 884

RESULT 7
O8VCK1 PRELIMINARY; PRT; 349 AA.
AC O8VCK1;
DT 01-MAR-2002 (TRENBLREL. 20, Created)
DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)
DT 01-MAR-2004 (TRENBLREL. 26, Last annotation update)
DE BC019693 protein.
GN Name=BC019693;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=mix FVB/N;
RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RX MEDLINE=22388257; PubMed=1477932;
RA Klausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stepleson M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein W.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fanej J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blackley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Skalska U., Schmutz J., Myers R.M., Buterfield Y.S.,
RA Kirylnski M.I., Skalska U., Smallos D.E., Schermer A., Schein J.E.,
RA Jones S.J., Maira M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=mix FVB/N;
RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RA Strauberg R.;
RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC019693; AAH19693.1;
DR MGD; MGI:2384983; BC019693.
DR InterPro: IPR008938; ARM.
```

```
SQL SEQUENCE 349 AA; 40166 MW; 9763D0331AD0F515 CRC64;

Query Match 34.1%; Score 14; DB 2; Length 349;
Best Local Similarity 100.0%; Pred. No. 7.2e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 18 PCIAQSFVAMADDS 31
DB 252 PCIAQSFVAMADDS 265

RESULT 8
O8CCT5 PRELIMINARY; PRT; 408 AA.
AC O8CCT5;
DT 01-MAR-2003 (TRENBLREL. 23, Created)
DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)
DT 01-MAR-2004 (TRENBLREL. 26, Last annotation update)
DE Mus musculus adult male olfactory brain cDNA, RIKEN full-length
DE enriched library, clone:643040D06 product:hypothetical ARM repeat
DE structure containing protein, full insert sequence.
GN Name=BC019693;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RC MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RC MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RC THE FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RC MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugihara Y., Shibata K., Itoh M.,
RA Konno H., Ozaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RC MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuji T., Taishiro H., Itoh M.,
RA Sumi N., Iehli Y., Nakamura S., Hazama M., Nishino T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kita A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multichannel sequencer.";
RN [6]
RP SEQUENCE FROM N.A.
```

RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiroaka T., Hirozane T.,
 RA Hoti F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
 RA Kato H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
 RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK032134; BAC27721.1; -
 DR MGI: 2384983; BC019693.
 DR InterPro: IPR008938; ARM.
 SQ HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 408 AA; 46688 MW; 42AB9EB13CA3FE67 CRC64;
 Query Match 34.1%; Score 14; DB 2; Length 408;
 Best Local Similarity 100.0%; Pred. No. 8.2e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 18 PCIAQSVYAMADS 31
 Db 311 PCIAQSVYAMADS 324

RESULT 9
 ID Q7XP06 PRELIMINARY; PRT; 97 AA.
 AC Q7XP06;
 DT 01-OCT-2003 (TREMBLrel. 25, Last Created)
 DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE OS=NB001313.14 protein.
 GN Name=OSNB001313.14;
 OS Oryza sativa (Japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriactoidae; Oryzae; Oryza.
 OC NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=12447439;
 RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,
 RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y.,
 RA Weng Q., Zhang L., Lu Y., Mu Y., Lu Y., Zhang L.S., Yu Z., Fan D.,
 RA Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,
 RA Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H.,
 RA Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,
 RA Chen J., Kang H., Chen C., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,
 RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,
 RA Lan L., Lei Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,
 RA Han B.;
 RT "Sequence and analysis of rice chromosome 4.";
 RL Nature 420:316-320(2002).
 DR EMBL: AL662939; CA03837.3; -
 DR Genbank: Q7XP06; -
 SQ SEQUENCE 97 AA; 10397 MW; DB293F2C47209DB6 CRC64;
 Query Match 17.1%; Score 7; DB 2; Length 97;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 26 AMADDSL 32
 Db 62 AMADDSL 68

RESULT 10
 ID Q9N045 PRELIMINARY; PRT; 285 AA.

AC Q9N045;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE Unnamed protein product.
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 OC NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21458551; PubMed=11574149;
 RA Oseida N., Hida M., Kusuda Y., Tanuma R., Iseki K., Hirata M., Suto Y.,
 RA Hirai M., Terao K., Suzuki Y., Sugano S., Hashimoto K.;
 RT "Assignment of 118 novel cDNAs of cynomolgus monkey brain to human
 RT chromosomes.";
 RL Gene 275:31-37(2001).
 DR EMBL: AB046087; BAB01669.1; -
 SQ SEQUENCE 285 AA; 31197 MW; BB4FC0ACC8261DB1 CRC64;
 Query Match 17.1%; Score 7; DB 2; Length 285;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 11 RVTKHLI 17
 Db 9 RVTKHLI 15

RESULT 11
 ID Q80VS5 PRELIMINARY; PRT; 365 AA.
 AC Q80VS5;
 DT 01-JUN-2003 (TREMBLrel. 24, Created)
 DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE Hypothetical protein (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=mx FVB/N;
 RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Dichtenko L., Marziani K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Umedin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
 RA Krzywinski M.I., Skelton U., Smalins D.E., Schermer A., Schein J.E.,
 RA Jones S.J., Warr M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=mx FVB/N;
 RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
 RA Strausberg R.;
 RA Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.


```

Db          211 GEEKFOE 217

RESULT 14
AAH70973    PRELIMINARY;      PRT;      625 AA.
AC          AAH70973;
DT 24-MAY-2004 (TREMBlrel. 27, Created)
DT 24-MAY-2004 (TREMBlrel. 27, Last sequence update)
DT 24-MAY-2004 (TREMBlrel. 27, Last annotation update)
DE Hypotheetical protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae;
OC Xenopodidae; Xenopus; Xenopus.
CX NCBI_TaxID=8335;
RN          [1]
RP          SEQUENCE FROM N.A.
RC          TISSUE=Embryo;
RX          MEDLINE=22341132; PubMed=12454917;
RA          Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA          Richardson P.;
RT          "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT          Initiative."
RL          Dev. Dyn. 225:384-391 (2002).
RN          [2]
RP          SEQUENCE FROM N.A.
RC          TISSUE=Embryo;
RX          MEDLINE=22388257; PubMed=12477932;
RA          Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA          Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA          Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA          Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA          Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA          Stedlecken M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA          Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA          Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA          Bosnak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA          Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huliyk S.W.,
RA          Villalón D.K., Mazny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA          Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA          Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
RA          Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA          Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.,
RA          Krzywnski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,
RA          Jones S.J., Marra M.A.;
RT          "Generation and initial analysis of more than 15,000 full-length human
RT          and mouse cDNA sequences."
RL          Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN          [3]
RP          SEQUENCE FROM N.A.
RC          TISSUE=Embryo;
RX          Klein S., Strausberg R.;
RL          Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR          EMBL; BC070973; AAH70973.1; -.
KW          Hypotheetical protein.
SO          SEQUENCE 625 AA; 70243 MW; 1C0947344FD3DAB CRC64;

Query Match          17.1%; Score 7; DB 2; Length 625;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          4 GEEKFOE 10
DB          211 GEEKFOE 217

```

```

RESULT 15
O29993      PRELIMINARY;      PRT;      639 AA.
AC          O29993;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

```

```

DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Iron (II) transporter (Feob-1).
GN OrderedLocustNames=AF0246;
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
CX NCBI_TaxID=2234;
RN          [1]
RP          SEQUENCE FROM N.A.
RC          STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX          MEDLINE=98049343; PubMed=9389475; DOI=10.1038/37052;
RA          Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA          Kechum K.A., Dodson R.J., Gwin M.L., Hickey E.K., Peterson J.D.,
RA          Richardson D.L., Quackenbush J., Lee N.H., Sutton G.G., Gill S.R.,
RA          Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.U.,
RA          Peterson S.N., Reich C.I., McNeil L.K., Badger J.H., Glodok A.,
RA          Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L.A.,
RA          Uteerback T.R., Cotton M.D., Spriggs T., Artlich P., Kaine B.P.,
RA          Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C.,
RA          Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O.,
RA          Moese C.R., Venter J.C.;
RT          "The complete genome sequence of the hyperthermophilic, sulphate-
RT          reducing archaeon Archaeoglobus fulgidus."
RL          Nature 390:364-370 (1997).
DR          EMBL; AE001089; AAB90987.1; -.
DR          PIR; F69280; F69280.
DR          HSP; O9XIF8; IMKY.
DR          TIGR; AF0246; -.
DR          GO; GO:0016020; C:membrane; IEA.
DR          GO; GO:0015093; F:ferrous iron transporter activity; IEA.
DR          GO; GO:0005525; F:GTP binding; IEA.
DR          GO; GO:0015684; P:ferrous iron transport; IEA.
DR          InterPro; IPR003573; FeOB.
DR          InterPro; IPR005289; GTP-binding.
DR          InterPro; IPR006073; GTP1_OBG.
DR          InterPro; IPR005225; Small_GTP.
DR          Pfam; PF02421; FeOB_1.
DR          PRINTS; PR00326; GTP1_OBG.
DR          TIGRFAMs; TIGR00437; FeOB; 1.
DR          TIGRFAMs; TIGR00650; MG442; 1.
DR          TIGRFAMs; TIGR00231; small_gtp; 1.
KW          Complete proteome.
SO          SEQUENCE 639 AA; 70367 MW; 6B39B71323C981FB CRC64;

Query Match          17.1%; Score 7; DB 2; Length 639;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          18 PCIAOFS 24
DB          597 PCIAOFS 603

```

Search completed: February 1, 2005, 15:31:09
 Job time : 72.5556 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:58:28 ; Search time 17.5556 Seconds
(without alignment)
154.882 Million cell updates/sec

Title: SEQ5_2030_2070

Perfect score: 41
Sequence: 1 RLGESEKFERVTKHLIPCI.....QFSVAMDDSLMKPLNYQIL 41

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size : 0

Total number of hits satisfying chosen parameters: 290433

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents, AA:*

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6C.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/6D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	17.1	98	1 US-08-211-202-130	Sequence 130, App
2	7	17.1	117	3 US-08-545-809A-139	Sequence 139, App
3	7	17.1	418	4 US-09-328-352-8089	Sequence 8089, App
4	6	14.6	113	4 US-09-733-210-774	Sequence 774, App
5	6	14.6	123	4 US-09-543-681A-8221	Sequence 8221, App
6	6	14.6	123	4 US-09-248-796A-17764	Sequence 17764, App
7	6	14.6	125	4 US-09-370-838-59	Sequence 838, App
8	6	14.6	125	4 US-09-854-133-59	Sequence 59, App1
9	6	14.6	137	4 US-09-489-039A-7848	Sequence 7848, App
10	6	14.6	145	4 US-09-270-767-36572	Sequence 36572, App
11	6	14.6	145	4 US-09-270-767-51789	Sequence 51789, App
12	6	14.6	149	4 US-09-252-991A-20842	Sequence 20842, App
13	6	14.6	154	3 US-09-247-155-110	Sequence 110, App
14	6	14.6	155	4 US-09-149-476-466	Sequence 466, App
15	6	14.6	155	4 US-09-149-476-603	Sequence 603, App
16	6	14.6	161	5 PCT-US95-05741-11	Sequence 11, App1
17	6	14.6	162	4 US-09-328-352-4393	Sequence 4393, App
18	6	14.6	222	4 US-09-328-352-8105	Sequence 8105, App
19	6	14.6	242	4 US-09-252-991A-17571	Sequence 17571, App
20	6	14.6	255	4 US-09-489-039A-10395	Sequence 10395, App
21	6	14.6	258	3 US-09-134-001C-2955	Sequence 2955, App
22	6	14.6	281	4 US-09-270-767-58539	Sequence 58539, App
23	6	14.6	293	4 US-09-270-767-43443	Sequence 43443, App
24	6	14.6	352	2 US-08-483-926A-11	Sequence 11, App1
25	6	14.6	352	2 US-08-737-045-12	Sequence 12, App1
26	6	14.6	387	4 US-09-489-039A-14027	Sequence 14027, App
27	6	14.6	405	4 US-09-252-991A-17862	Sequence 17862, App

28	6	14.6	426	4 US-09-252-991A-10025	Sequence 30025, App
29	6	14.6	452	4 US-09-889-738-21	Sequence 21, App1
30	6	14.6	457	4 US-09-721-870-26	Sequence 26, App1
31	6	14.6	480	4 US-09-252-991A-26186	Sequence 26186, App
32	6	14.6	508	4 US-09-270-767-43201	Sequence 43201, App
33	6	14.6	528	4 US-09-252-991A-26090	Sequence 26090, App
34	6	14.6	601	4 US-09-252-991A-18877	Sequence 18877, App
35	6	14.6	649	4 US-09-418-963-2	Sequence 2, App1
36	6	14.6	744	4 US-09-252-991A-19290	Sequence 19290, App
37	6	14.6	790	4 US-09-252-991A-23247	Sequence 23247, App
38	6	14.6	896	4 US-09-543-681A-7870	Sequence 7870, App
39	6	14.6	941	4 US-09-513-783A-172	Sequence 172, App
40	6	14.6	960	3 US-09-345-650-1	Sequence 1, App1
41	6	14.6	998	4 US-09-540-236-3801	Sequence 3801, App
42	6	14.6	1584	3 US-09-457-040B-27	Sequence 27, App1
43	6	14.6	5588	3 US-09-036-987A-6	Sequence 6, App1
44	6	14.6	5588	3 US-09-370-700-6	Sequence 6, App1
45	6	14.6	5588	4 US-09-603-207-6	Sequence 6, App1

ALIGNMENTS

RESULT 1
US-08-211-202-130
Sequence 130, Application US/08211202
Patent No. 5565332
GENERAL INFORMATION:
APPLICANT: HOOGENBOOM, Hendricus Reneus Jacobus Matheus
APPLICANT: BAIER, Michael
APPLICANT: JESPEERS, Laurent Stephane Anne Therese
APPLICANT: WINTER, Gregory Paul
TITLE OF INVENTION: Production of chimeric antibodies - a
NUMBER OF SEQUENCES: 144
CORRESPONDENCE ADDRESSES:
ADDRESSEE: David W. Clough, Marshall O'Toole Gerstein Murray &
ADDRESS: Boston
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/211,202
FILING DATE: 23-SEP-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9120252.3
FILING DATE: 23-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9120377.8
FILING DATE: 25-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9206318.9
FILING DATE: 24-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9206372.6
FILING DATE: 24-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB92/00883
FILING DATE: 15-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: David W. Clough
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 28111/31960
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300

TELEFAX: 312-474-0448
TELEX: 25-3856 ID NO: 130:
INFORMATION FOR SEQ ID NO:
SEQUENCE CHARACTERISTICS:
LENGTH: 98 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-211-202-130

Query Match 17.1%; Score 7; DB 1; Length 98;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 KFOERYT 13
|||
Db 63 KFOERYT 69

RESULT 2
US-08-545-809A-139
Sequence 139, Application US/08545809A
Patent No. 6096878
GENERAL INFORMATION:
APPLICANT: Honjo, Tasuku
APPLICANT: Matsuda, Fumihiko
TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
NUMBER OF SEQUENCES: 145
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545.809A
FILING DATE: 27-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/00603
FILING DATE: 10-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 06501/004001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 139:
SEQUENCE CHARACTERISTICS:
LENGTH: 117 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-545-809A-139

Query Match 17.1%; Score 7; DB 3; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 KFOERYT 13
|||
Db 82 KFOERYT 88

RESULT 3

US-09-328-352-8089
Sequence 8089, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Bretton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328.352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 8089
LENGTH: 418
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-8089

Query Match 17.1%; Score 7; DB 4; Length 418;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 IAOBSVA 26
|||
Db 22 IAOBSVA 28

RESULT 4
US-09-732-210-774
Sequence 774, Application US/09732210
Patent No. 6573561
GENERAL INFORMATION:
APPLICANT: Bunkers, Greg J.
APPLICANT: Liang, Jihong
APPLICANT: Miltanck, Cindy A.
APPLICANT: Seale, Jeffrey W.
APPLICANT: Wu, Yomale S.
TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use
FILE REFERENCE: 38-21(15036)B
CURRENT APPLICATION NUMBER: US/09/732.210
CURRENT FILING DATE: 2000-12-07
PRIOR APPLICATION NUMBER: US 60/169,513
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: US 60/169,340
PRIOR FILING DATE: 1999-12-07
NUMBER OF SEQ ID NOS: 1753
SEQ ID NO 774
LENGTH: 113
TYPE: PRT
ORGANISM: Micrococcus luteus
US-09-732-210-774

Query Match 14.6%; Score 6; DB 4; Length 113;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RVTXHL 16
|||
Db 41 RVTXHL 46

RESULT 5
US-09-543-681A-8221
Sequence 8221, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543.681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09

NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 8221
LENGTH: 123
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-8221

Query Match 14.6%; Score 6; DB 4; Length 123;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 FOERTV 13
|||
Db 101 FOERTV 106

RESULT 6
US-09-248-796A-17764
Sequence 17764, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
PRIOR FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 17764
LENGTH: 123
TYPE: PRT
ORGANISM: Candida albicans
US-09-248-796A-17764

Query Match 14.6%; Score 6; DB 4; Length 123;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 TKHLIP 18
|||
Db 48 TKHLIP 53

RESULT 7
US-09-370-838-59
Sequence 59, Application US/09370838
Patent No. 6444425
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Mohamath, Rashed
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
FILE REFERENCE: 210121.475C1
CURRENT APPLICATION NUMBER: US/09/370,838
PRIOR FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/285,323
PRIOR FILING DATE: 1999-04-02
NUMBER OF SEQ ID NOS: 289
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 59
LENGTH: 125
TYPE: PRT
ORGANISM: Homo sapien
US-09-370-838-59

Query Match 14.6%; Score 6; DB 4; Length 125;
Best Local Similarity 100.0%; Pred. No. 24;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 20 IAOFSV 25
|||
Db 43 IAOFSV 48

RESULT 8
US-09-854-133-59
Sequence 59, Application US/09854133
Patent No. 6759508
GENERAL INFORMATION:
APPLICANT: Lodes, Michael J.
APPLICANT: Mohamath, Rashed
APPLICANT: Henderson, Robert A.
APPLICANT: Benson, Darin R.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
FILE REFERENCE: 210121.475C10
CURRENT APPLICATION NUMBER: US/09/854,133
PRIOR FILING DATE: 2001-05-11
NUMBER OF SEQ ID NOS: 735
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 59
LENGTH: 125
TYPE: PRT
ORGANISM: Homo sapien
US-09-854-133-59

Query Match 14.6%; Score 6; DB 4; Length 125;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 IAOFSV 25
|||
Db 43 IAOFSV 48

RESULT 9
US-09-489-039A-7848
Sequence 7848, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
PRIOR FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 7848
LENGTH: 137
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7848

Query Match 14.6%; Score 6; DB 4; Length 137;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 LMKPLN 37
|||
Db 104 LMKPLN 109

RESULT 10
US-09-270-767-36572
Sequence 36572, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:

APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO: 36572
LENGTH: 145
TYPE: PRT
ORGANISM: Drosophila melanogaster
FEATURE:
OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-36572

Query Match
Best Local Similarity 14.6%; Score 6; DB 4; Length 145;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 LNYQIL 41
DB 100 LNYQIL 105

RESULT 11
US-09-270-767-51789
Sequence 51789, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO: 51789
LENGTH: 145
TYPE: PRT
ORGANISM: Drosophila melanogaster
FEATURE:
OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-51789

Query Match
Best Local Similarity 14.6%; Score 6; DB 4; Length 145;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 LNYQIL 41
DB 100 LNYQIL 105

RESULT 12
US-09-252-991A-20842
Sequence 20842, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO: 20842
LENGTH: 149
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-20842

Query Match
Best Local Similarity 14.6%; Score 6; DB 4; Length 149;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 FSVAMA 28
DB 76 FSVAMA 81

RESULT 13
US-09-247-155-110
Sequence 110, Application US/09247155A
Patent No. 6312922
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, Jean-Baptiste
APPLICANT: Bouqueloret, Lydie
TITLE OF INVENTION: Complementary DNAs
FILE REFERENCE: GENSET 021A
CURRENT APPLICATION NUMBER: US/09/247,155A
CURRENT FILING DATE: 1999-02-09
EARLIER APPLICATION NUMBER: 60/074,121
EARLIER FILING DATE: 1998-02-09
EARLIER APPLICATION NUMBER: 60/081,563
EARLIER FILING DATE: 1998-04-13
EARLIER APPLICATION NUMBER: 60/096,116
EARLIER FILING DATE: 1998-08-10
EARLIER APPLICATION NUMBER: 60/099,273
EARLIER FILING DATE: 1998-10-04
NUMBER OF SEQ ID NOS: 182
SOFTWARE: Patent.pm
SEQ ID NO: 110
LENGTH: 154
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SIGNAL
LOCATION: -13..-1
US-09-247-155-110

Query Match
Best Local Similarity 14.6%; Score 6; DB 3; Length 154;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 PLNYQI 40
DB 58 PLNYQI 63

RESULT 14
US-09-149-476-466
Sequence 466, Application US/09149476
Patent No. 6420526
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 186 Human Secreted proteins
FILE REFERENCE: P2002P1
CURRENT APPLICATION NUMBER: US/09/149,476
CURRENT FILING DATE: 1998-09-08
EARLIER APPLICATION NUMBER: PCT/US98/04493
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040,162
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,334
EARLIER FILING DATE: 1997-03-07

EARLIER	FILING DATE:	1997-06-06
EARLIER	APPLICATION NUMBER:	60/056,886
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,877
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,889
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,893
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,630
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,878
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,882
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,662
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,872
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,903
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,888
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,879
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,880
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,636
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,874
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,910
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,864
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,631
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,845
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,892
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/057,761
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/047,595
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,599
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,588
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,585
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,586
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,590
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,594
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,589
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,593
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,614
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/043,578
EARLIER	FILING DATE:	1997-04-11
EARLIER	APPLICATION NUMBER:	60/043,576
EARLIER	FILING DATE:	1997-04-11

EARLIER APPLICATION NUMBER: 60/047,501
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,670
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/056,632
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,664
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,876
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,881
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,909
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,875
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,862
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,887
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,908
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/057,650
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/056,884
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,669
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/049,610
EARLIER FILING DATE: 1997-06-13
EARLIER APPLICATION NUMBER: 60/061,060
EARLIER FILING DATE: 1997-10-02

Query Match 14.6%; Score 6; DB 4;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 PLNYOI 40
DB 58 PLNYOI 63

RESULT 15
US-09-149-476-603
Sequence 603, Application US/09149476
Patent No. 6420526
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 186 Human Secreted proteins
FILE REFERENCE: P2002P1
CURRENT FILING DATE: US/09/149,476
EARLIER APPLICATION NUMBER: PCT/US98/04493
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040,162
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,334
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,336
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,163
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/047,600
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,615

EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,597
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,502
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,633
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,583
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,617
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,618
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,503
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,592
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,581
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,584
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,500
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,587
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,492
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,598
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,613
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,582
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,596
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,612
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,632
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,601
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,580
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,568
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,314
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,569
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,311
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,671
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,674
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,669
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,312
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,313
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,672
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,315
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/048,974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/056,886
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,877
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,889
EARLIER FILING DATE: 1997-08-22

EARLIER APPLICATION NUMBER: 60/056, 893
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 630
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 878
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 662
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 872
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 882
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 637
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 903
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 888
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 879
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 880
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 894
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 911
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 636
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 874
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 910
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 864
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 631
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 845
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 892
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/047, 761
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/047, 595
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 588
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 585
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 586
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 590
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 594
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 589
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 593
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043, 576
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/047, 614
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043, 578
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043, 576
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/047, 501
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043, 670
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/056, 632
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 664

EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 876
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 881
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 909
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 875
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 862
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 887
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 908
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/048, 964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/057, 650
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/056, 884
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057, 669
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/049, 610
EARLIER FILING DATE: 1997-06-13
EARLIER APPLICATION NUMBER: 60/061, 060
EARLIER FILING DATE: 1997-10-02

Query Match 14.6%; Score 6; DB 4; Length 155;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 PLNTQI 40
|||||
Db 58 PLNTQI 63

Search completed: February 1, 2005, 15:35:47
Job time : 18.5556 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 15:09:39 ; Search time 59.7778 Seconds
(without alignments)
247.799 Million cell updates/sec

Title: SEQ5_2030_2070

Perfect score: 41
Sequence: 1 RLGESEKFEQERVTGHLIPCI.....QFSVAMADSLMKLNYQIL 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1608061 seqs, 361289386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1311380

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published Applications_AA:*

1:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
12:	/cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13:	/cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
17:	/cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
18:	/cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
19:	/cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
20:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	1149	US-10-128-558-167	Sequence 167, App
2	8	19.5	132	US-10-424-599-200002	Sequence 200002,
3	8	19.5	188	US-10-424-599-199301	Sequence 199301,
4	7	17.1	36	US-09-864-761-44493	Sequence 44493, A
5	7	17.1	98	US-10-194-975-8	Sequence 8, Appl1
6	7	17.1	98	US-10-308-817-48	Sequence 8, Appl1
7	7	17.1	98	US-10-032-0378-38	Sequence 38, Appl1
8	7	17.1	98	US-10-029-9888-38	Sequence 38, Appl1
9	7	17.1	98	US-10-032-423A-38	Sequence 38, Appl1
10	7	17.1	98	US-10-453-698-48	Sequence 38, Appl1
11	7	17.1	98	US-10-029-926B-38	Sequence 38, Appl1
12	7	17.1	98	US-10-379-392-8	Sequence 8, Appl1
13	7	17.1	238	US-10-767-701-41775	Sequence 41775, A

14	7	17.1	600	US-10-112-944-472	Sequence 472, App
15	6	14.6	47	US-10-425-115-187478	Sequence 187478,
16	6	14.6	61	US-10-424-599-245649	Sequence 245649,
17	6	14.6	72	US-10-437-963-178932	Sequence 178932,
18	6	14.6	74	US-10-425-115-191570	Sequence 191570,
19	6	14.6	83	US-10-425-115-295263	Sequence 295263,
20	6	14.6	90	US-10-424-599-190438	Sequence 190438,
21	6	14.6	96	US-10-156-761-14128	Sequence 14128, A
22	6	14.6	97	US-10-437-963-180730	Sequence 180730,
23	6	14.6	97	US-10-767-701-52426	Sequence 52426, A
24	6	14.6	102	US-10-425-115-200993	Sequence 200993,
25	6	14.6	112	US-10-425-115-219369	Sequence 219369,
26	6	14.6	118	US-10-437-963-148111	Sequence 148111,
27	6	14.6	118	US-10-425-115-214549	Sequence 214549,
28	6	14.6	119	US-10-222-723-12	Sequence 12, Appl1
29	6	14.6	120	US-10-437-963-173400	Sequence 173400,
30	6	14.6	123	US-10-767-701-41325	Sequence 41325, A
31	6	14.6	125	US-09-738-973-59	Sequence 59, Appl1
32	6	14.6	125	US-09-854-133-59	Sequence 59, Appl1
33	6	14.6	125	US-10-144-649A-59	Sequence 59, Appl1
34	6	14.6	129	US-10-425-115-345168	Sequence 345168,
35	6	14.6	138	US-10-424-599-280443	Sequence 280443,
36	6	14.6	140	US-10-437-963-189779	Sequence 189779,
37	6	14.6	141	US-09-734-569-108	Sequence 108, App
38	6	14.6	141	US-10-437-963-102617	Sequence 102617,
39	6	14.6	148	US-10-437-963-204095	Sequence 204095,
40	6	14.6	154	US-09-903-190-110	Sequence 110, App
41	6	14.6	154	US-09-978-360A-766	Sequence 766, App
42	6	14.6	155	US-09-809-391-466	Sequence 466, App
43	6	14.6	155	US-09-809-391-603	Sequence 603, App
44	6	14.6	155	US-09-882-171-466	Sequence 466, App
45	6	14.6	155	US-09-882-171-603	Sequence 603, App

ALIGNMENTS

RESULT 1
US-10-128-558-167
Sequence 167, Application US/10128558
Publication No. US20040219521A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Wang, Zhilwei
APPLICANT: Weng, Gezhi
APPLICANT: Boyle, Bryan J
APPLICANT: Drmanac, Radcoje T
TITLE OF INVENTION: Novel Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 812A
CURRENT APPLICATION NUMBER: US/10/128, 558
CURRENT FILING DATE: 2002-04-22
PRIOR APPLICATION NUMBER: US 60/339, 453
PRIOR FILING DATE: 2001-12-11
PRIOR APPLICATION NUMBER: US 09/488, 725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552, 317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491, 404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02663
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496, 914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/560, 875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: PCT/US01/03800
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 09/515, 126
PRIOR FILING DATE: 2000-02-28
Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 412
SOFTWARE: PE_FL_genes Version 6.0
SEQ ID NO 167
LENGTH: 1149
TYPE: PRT
ORGANISM: Homo sapiens
US-10-128-558-167

Query Match 100.0%; Score 41; DB 17; Length 1149;
Best Local Similarity 100.0%; Pred. No. 5.4e-35;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RLGGEKFEQERYTKALIPICIAQSVAMADSLMKPLNYOIL 41
Db 1035 RLGGEKFEQERYTKALIPICIAQSVAMADSLMKPLNYOIL 1075

RESULT 2
US-10-424-599-200002
Sequence 200002, Application US/10424599
Publication No. US20040031072A1

GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 200002
LENGTH: 132
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_22628C.1.pep
US-10-424-599-200002

Query Match 19.5%; Score 8; DB 15; Length 132;
Best Local Similarity 100.0%; Pred. No. 0.96;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 VAMADSL 32
Db 9 VAMADSL 16

RESULT 3
US-10-424-599-199301
Sequence 199301, Application US/10424599
Publication No. US20040031072A1

GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 199301
LENGTH: 188
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_219930C.1.pep
US-10-424-599-199301

Query Match 19.5%; Score 8; DB 15; Length 188;

Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 VAMADSL 32
Db 14 VAMADSL 21

RESULT 4
US-09-864-761-44493
Sequence 44493, Application US/09864761
Patent No. US20020048763A1

GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 44493
LENGTH: 36
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005229.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.96
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.9
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.86
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.86
US-09-864-761-44493

Query Match 17.1%; Score 7; DB 9; Length 36;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGEERFQ 9
|||||
DB 4 GGEERFQ 10

RESULT 5
US-10-194-975-8
; Sequence 8, Application US/10194975
; Publication No. US2003039649A1
; GENERAL INFORMATION:
; APPLICANT: Foote, Jefferson
; TITLE OF INVENTION: Super Humanized Antibodies
; FILE REFERENCE: 501231.01
; CURRENT APPLICATION NUMBER: US/10/194,975
; CURRENT FILING DATE: 2002-10-10
; PRIOR APPLICATION NUMBER: US 60/305,111
; PRIOR FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-194-975-8

Query Match 17.1%; Score 7; DB 14; Length 98;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 KFOERVT 13
|||||
DB 63 KFOERVT 69

RESULT 6
US-10-308-817-48
; Sequence 48, Application US/10308817
; Publication No. US20030219861A1
; GENERAL INFORMATION:
; APPLICANT: Rothen, Russell
; APPLICANT: Wu, Dayang
; TITLE OF INVENTION: HYBRID ANTIBODIES
; FILE REFERENCE: 1087-37
; CURRENT APPLICATION NUMBER: US/10/308,817
; CURRENT FILING DATE: 2002-12-03
; NUMBER OF SEQ ID NOS: 195
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48
; LENGTH: 98
; TYPE: PRT
; ORGANISM: human
US-10-308-817-48

Query Match 17.1%; Score 7; DB 14; Length 98;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 KFOERVT 13
|||||
DB 63 KFOERVT 69

RESULT 7
US-10-032-037B-38
; Sequence 38, Application US/10032037B
; Publication No. US20040001822A1
; GENERAL INFORMATION:
; APPLICANT: Bio-Technology General Corp.

; TITLE OF INVENTION: Y17-ISOLATED MOLECULES COMPRISING EPITOPES CONTAINING SULFATED
; FILE REFERENCE: 10793/44
; CURRENT APPLICATION NUMBER: US/10/032,037B
; CURRENT FILING DATE: 2001-12-31
; PRIOR APPLICATION NUMBER: 60/258,948
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 38
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-037B-38

Query Match 17.1%; Score 7; DB 15; Length 98;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 KFOERVT 13
|||||
DB 63 KFOERVT 69

RESULT 8
US-10-029-988B-38
; Sequence 38, Application US/10029988B
; Publication No. US20040001839A1
; GENERAL INFORMATION:
; APPLICANT: Bio-Technology General Corp.
; TITLE OF INVENTION: Y17-ISOLATED MOLECULES COMPRISING EPITOPES CONTAINING SULFATED
; FILE REFERENCE: 10793/46
; CURRENT APPLICATION NUMBER: US/10/029,988B
; CURRENT FILING DATE: 2001-12-31
; PRIOR APPLICATION NUMBER: 60/258,948
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 38
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-029-988B-38

Query Match 17.1%; Score 7; DB 15; Length 98;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 KFOERVT 13
|||||
DB 63 KFOERVT 69

RESULT 9
US-10-032-423A-38
; Sequence 38, Application US/10032423A
; Publication No. US20040002450A1
; GENERAL INFORMATION:
; APPLICANT: Bio-Technology General Corp.
; TITLE OF INVENTION: Y17-ISOLATED MOLECULES COMPRISING EPITOPES CONTAINING SULFATED
; FILE REFERENCE: 10793/45
; CURRENT APPLICATION NUMBER: US/10/032,423A
; CURRENT FILING DATE: 2001-12-31
; PRIOR APPLICATION NUMBER: 60/258,948
; PRIOR FILING DATE: 12/29/2000
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 38
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Homo sapiens

US-10-032-423A-38

Query Match 17.1%; Score 7; DB 15; Length 98;
 Best Local Similarity 100.0%; Pred. No. 9.2;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 KFOERVT 13
 |||||
 Db 63 KFOERVT 69

RESULT 10

US-10-453-698-48
 ; Sequence 48, Application US/10453698
 ; Publication No. US20040038308A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rother, Russell
 ; TITLE OF INVENTION: HYBRID ANTIBODIES
 ; FILE REFERENCE: 82 CIP (1087-37 CIP)
 ; CURRENT APPLICATION NUMBER: US/10/453,698
 ; CURRENT FILING DATE: 2003-06-03
 ; NUMBER OF SEQ ID NOS: 196
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 48
 ; LENGTH: 98
 ; TYPE: PRT
 ; ORGANISM: human
 US-10-453-698-48

Query Match 17.1%; Score 7; DB 15; Length 98;
 Best Local Similarity 100.0%; Pred. No. 9.2;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 KFOERVT 13
 |||||
 Db 63 KFOERVT 69

RESULT 11

US-10-029-926B-38
 ; Sequence 38, Application US/10029926B
 ; Publication No. US20040073011A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HAGAY, et al.
 ; TITLE OF INVENTION: SPECIFIC HUMAN ANTIBODIES FOR SELECTIVE CANCER THERAPY
 ; FILE REFERENCE: 10793/50
 ; CURRENT APPLICATION NUMBER: US/10/029,926B
 ; CURRENT FILING DATE: 2001-12-31
 ; PRIOR APPLICATION NUMBER: 60/258,948
 ; PRIOR FILING DATE: 12/29/2000
 ; NUMBER OF SEQ ID NOS: 203
 ; SOFTWARE: FastSeq For Windows Version 3.0
 ; SEQ ID NO 38
 ; LENGTH: 98
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-029-926B-38

Query Match 17.1%; Score 7; DB 15; Length 98;
 Best Local Similarity 100.0%; Pred. No. 9.2;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 KFOERVT 13
 |||||
 Db 63 KFOERVT 69

RESULT 12

US-10-379-392-8
 ; Sequence 8, Application US/10379392
 ; Publication No. US20040110226A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lazar, Gregory Alan

; APPLICANT: Desjarlais, John Rudolf
 ; APPLICANT: Marshall, Shannon Alicia
 ; APPLICANT: Dahiyat, Basil I.
 ; TITLE OF INVENTION: ANTIBODY OPTIMIZATION
 ; FILE REFERENCE: A-71386-3 463077-236
 ; CURRENT APPLICATION NUMBER: US/10/379,392
 ; CURRENT FILING DATE: 2003-03-03
 ; PRIOR APPLICATION NUMBER: US 60/360,843
 ; PRIOR FILING DATE: 2002-03-01
 ; PRIOR APPLICATION NUMBER: US 60/384,197
 ; PRIOR FILING DATE: 2002-05-29
 ; NUMBER OF SEQ ID NOS: 184
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 8
 ; LENGTH: 98
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-379-392-8

Query Match 17.1%; Score 7; DB 16; Length 98;
 Best Local Similarity 100.0%; Pred. No. 9.2;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 KFOERVT 13
 |||||
 Db 63 KFOERVT 69

RESULT 13

US-10-767-701-41775
 ; Sequence 41775, Application US/107677701
 ; Publication No. US20040172684A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Zhou, Yihua
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
 ; FILE REFERENCE: 38-21(53535)B
 ; CURRENT APPLICATION NUMBER: US/10/767,701
 ; CURRENT FILING DATE: 2004-01-29
 ; NUMBER OF SEQ ID NOS: 63128
 ; SEQ ID NO 41775
 ; LENGTH: 238
 ; TYPE: PRT
 ; ORGANISM: Sorghum bicolor
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C6863_1.pep
 US-10-767-701-41775

Query Match 17.1%; Score 7; DB 16; Length 238;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 LGGEKRF 8
 |||||
 Db 211 LGGEKRF 217

RESULT 14

US-10-112-944-472
 ; Sequence 472, Application US/10112944
 ; Publication No. US20040048249A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Y. Tom
 ; APPLICANT: Yang, Yonghong
 ; APPLICANT: Weng, Gezh
 ; APPLICANT: Zhang, Jie
 ; APPLICANT: Ren, Feiyan
 ; APPLICANT: Xue, Aidong J.
 ; APPLICANT: Wang, Jian-Rui
 ; APPLICANT: Wehrman, Tom
 ; APPLICANT: Ghosh, Malabika

Search completed: February 1, 2005, 15:44:53
 Job time : 60.7778 secs

```

; APPLICANT: Wang, Dunrui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Zhiwei
; TITLE OF INVENTION: No. US20040048249A1el Nucleic Acids and
; FILE REFERENCE: 805A
; CURRENT APPLICATION NUMBER: US/10/112,944
; FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: US 09/519,705
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: US 09/552,929
; PRIOR FILING DATE: 2000-04-16
; PRIOR APPLICATION NUMBER: US 09/577,408
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 924
; SOFTWARE: pf_fl_genes Version 5.0
; SEQ ID NO 472
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-112-944-472

```

```

Query Match      17.1%; Score 7; DB 15; Length 600;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      7 KFORVTV 13
      |||||
Db      3 KFORVTV 9

```

```

RESULT 15
US-10-425-115-187478
; Sequence 187478, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 187478
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_102564C.1.pep
US-10-425-115-187478

```

```

Query Match      14.6%; Score 6; DB 17; Length 47;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      8 FORVTV 13
      |||||
Db      26 FORVTV 31

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:58:28 ; Search time 17.5556 Seconds
(without alignments)
154.882 Million cell updates/sec

Title: SEQ5A1A1854

Perfect score: 41
Sequence: 1 EKNMKNHMGPFMSILOEHIG.....MKKEELTSHQSQTAFLEA 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size : 0

Total number of hits satisfying chosen parameters: 230433

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.dep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.dep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.dep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.dep:*
5: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.dep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.dep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	19.5	411	US-09-489-039A-7912	Sequence 7912, Ap
2	8	19.5	457	US-09-543-681A-6481	Sequence 6481, Ap
3	8	19.5	473	US-09-489-039A-9854	Sequence 9854, Ap
4	8	19.5	480	US-09-543-681A-4532	Sequence 4532, Ap
5	8	19.5	492	US-09-252-981A-20403	Sequence 20403, A
6	7	17.1	431	US-09-286-981B-3	Sequence 3, Appl1
7	7	17.1	588	US-08-714-741-42	Sequence 42, Appl
8	7	17.1	605	US-08-714-741-46	Sequence 46, Appl
9	7	17.1	864	US-08-714-741-40	Sequence 40, Appl
10	7	17.1	1231	US-08-714-741-41	Sequence 41, Appl
11	6	14.6	51	US-09-513-999C-4343	Sequence 4343, Ap
12	6	14.6	56	US-09-270-767-57083	Sequence 57083, A
13	6	14.6	97	US-09-621-976-5794	Sequence 5794, Ap
14	6	14.6	142	US-09-270-767-41841	Sequence 41841, A
15	6	14.6	143	US-09-198-452A-6	Sequence 6, Appl1
16	6	14.6	162	US-09-134-001C-3670	Sequence 3670, Ap
17	6	14.6	226	US-09-176-657-3	Sequence 3, Appl1
18	6	14.6	226	US-09-421-239-3	Sequence 3, Appl1
19	6	14.6	230	US-08-540-118-1	Sequence 1, Appl1
20	6	14.6	230	US-09-185-818-1	Sequence 6, Appl1
21	6	14.6	232	US-09-328-352-6245	Sequence 6245, A
22	6	14.6	254	US-09-489-039A-12987	Sequence 12987, A
23	6	14.6	263	US-08-845-258-18	Sequence 18, Appl
24	6	14.6	263	US-08-990-571-18	Sequence 18, Appl
25	6	14.6	263	US-08-723-142A-18	Sequence 18, Appl
26	6	14.6	263	US-09-528-784A-18	Sequence 18, Appl
27	6	14.6	263	US-09-569-098A-18	Sequence 18, Appl

28	6	14.6	267	US-08-845-258-28	Sequence 28, Appl
29	6	14.6	267	US-08-990-571-28	Sequence 28, Appl
30	6	14.6	267	US-08-723-142A-28	Sequence 28, Appl
31	6	14.6	267	US-09-528-784A-28	Sequence 28, Appl
32	6	14.6	267	US-09-569-098A-28	Sequence 28, Appl
33	6	14.6	294	US-08-845-258-46	Sequence 46, Appl
34	6	14.6	294	US-08-990-571-46	Sequence 46, Appl
35	6	14.6	294	US-08-723-142A-46	Sequence 46, Appl
36	6	14.6	294	US-09-528-784A-46	Sequence 46, Appl
37	6	14.6	294	US-09-569-098A-46	Sequence 46, Appl
38	6	14.6	303	US-08-845-258-23	Sequence 23, Appl
39	6	14.6	303	US-08-990-571-23	Sequence 23, Appl
40	6	14.6	303	US-08-723-142A-23	Sequence 23, Appl
41	6	14.6	303	US-09-528-784A-23	Sequence 23, Appl
42	6	14.6	310	US-09-569-098A-23	Sequence 23, Appl
43	6	14.6	310	US-08-845-258-19	Sequence 19, Appl
44	6	14.6	310	US-08-990-571-19	Sequence 19, Appl
45	6	14.6	310	US-08-723-142A-19	Sequence 19, Appl

ALIGNMENTS

```
RESULT 1
US-09-489-039A-7912
: Sequence 7912, Application US/09489039A
: Patent No. 6610836
: GENERAL INFORMATION:
: APPLICANT: Gary Breton et. al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
: FILE REFERENCE: 2709.2004001
: CURRENT APPLICATION NUMBER: US/09/489, 039A
: PRIOR FILING DATE: 2000-01-27
: PRIOR APPLICATION NUMBER: US 60/117,747
: NUMBER OF SEQ ID NOS: 14342
: SEQ ID NO 7912
: LENGTH: 411
: TYPE: PRT
: ORGANISM: Klebsiella pneumoniae
: US-09-489-039A-7912

Query Match          19.5%; Score 8; DB 4; Length 411;
Best Local Similarity 100.0%; Pred. No. 0.49;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      34 LTAFFLEA 41
      |||||
Db      46 LTAFFLEA 53

RESULT 2
US-09-543-681A-6481
: Sequence 6481, Application US/09543681A
: Patent No. 6605709
: GENERAL INFORMATION:
: APPLICANT: GARY BRETON
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
: FILE REFERENCE: 2709.1002-001
: CURRENT APPLICATION NUMBER: US/09/543, 681A
: PRIOR FILING DATE: 2000-04-05
: PRIOR APPLICATION NUMBER: US 60/128,706
: NUMBER OF SEQ ID NOS: 8344
: SEQ ID NO 6481
: LENGTH: 457
: TYPE: PRT
: ORGANISM: Proteus mirabilis
: US-09-543-681A-6481

Query Match          19.5%; Score 8; DB 4; Length 457;
```

Best Local Similarity 100.0%; Pred. No. 0.53;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 LTAFFLEA 41
Db 110 LTAFFLEA 117

RESULT 3

US-09-489-039A-9854
Sequence 9854; Application US/09489039A

Patent No. 6610836
GENERAL INFORMATION:

APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A

PRIOR FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747

NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 9854

LENGTH: 473
TYPE: PRT

ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9854

Query Match

Best Local Similarity 19.5%; Score 8; DB 4; Length 473;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 LTAFFLEA 41
Db 111 LTAFFLEA 118

RESULT 4

US-09-543-681A-4532
Sequence 4532; Application US/09543681A

Patent No. 6605709
GENERAL INFORMATION:

APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A

PRIOR FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706

PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344

SEQ ID NO 4532
LENGTH: 480

TYPE: PRT
ORGANISM: Proteus mirabilis

US-09-543-681A-4532

Query Match
Best Local Similarity 19.5%; Score 8; DB 4; Length 480;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 LTAFFLEA 41
Db 139 LTAFFLEA 146

RESULT 5

US-09-252-991A-20403
Sequence 20403; Application US/09252991A

Patent No. 6551795
GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 20403
LENGTH: 492

TYPE: PRT
ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-20403

Query Match
Best Local Similarity 19.5%; Score 8; DB 4; Length 492;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 LTAFFLEA 41
Db 107 LTAFFLEA 114

RESULT 6

US-09-286-981B-3
Sequence 3; Application US/09286981B

Patent No. 6503511
GENERAL INFORMATION:

APPLICANT: Wismann, Theresa M.
APPLICANT: Koenig, Scott

APPLICANT: Johnson, Leslie S
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines

FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/09/286,981B

PRIOR FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743

PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38

SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3

LENGTH: 431
TYPE: PRT

ORGANISM: Streptococcus pneumoniae
US-09-286-981B-3

Query Match
Best Local Similarity 17.1%; Score 7; DB 4; Length 431;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 KKEBELTS 29
Db 74 KKEBELTS 80

RESULT 7

US-08-714-741-42
Sequence 42; Application US/08714741

Patent No. 6500613
GENERAL INFORMATION:

APPLICANT: Briles, David E.
APPLICANT: McDaniel, Larry S.

APPLICANT: Swiatlo, Edwin
APPLICANT: Yoher, Janet

APPLICANT: Crain, Marilyn J.
APPLICANT: Hollingshead, Susan

APPLICANT: Tart, Rebecca
APPLICANT: Brooks-Walter, Alexis

TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:

ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,741
FILING DATE: 16-SEP-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Frommer Esq., William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454312-2460
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 588 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: amino acid
US-08-714-741-42

Query Match 17.1%; Score 7; DB 4; Length 588;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 23 KKEELTS 29
Db 323 KKEELTS 329

RESULT 8
US-08-714-741-46
Sequence 46, Application US/08714741
Patent No. 650613
GENERAL INFORMATION:
APPLICANT: Briles, David E.
APPLICANT: McDaniel, Larry S.
APPLICANT: Swiatlo, Edwin
APPLICANT: Yother, Janet
APPLICANT: Crain, Marilyn J.
APPLICANT: Hollingshead, Susan
APPLICANT: Tart, Rebecca
APPLICANT: Brooks-Walter, Alexis
TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
TITLE OF INVENTION: PORTIONS AND PRODUCTS
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,741
FILING DATE: 16-SEP-1996
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Frommer Esq., William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454312-2460
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 605 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: amino acid
US-08-714-741-46

Query Match 17.1%; Score 7; DB 4; Length 605;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 23 KKEELTS 29
Db 106 KKEELTS 112

RESULT 9
US-08-714-741-40
Sequence 40, Application US/08714741
Patent No. 650613
GENERAL INFORMATION:
APPLICANT: Briles, David E.
APPLICANT: McDaniel, Larry S.
APPLICANT: Swiatlo, Edwin
APPLICANT: Yother, Janet
APPLICANT: Crain, Marilyn J.
APPLICANT: Hollingshead, Susan
APPLICANT: Tart, Rebecca
APPLICANT: Brooks-Walter, Alexis
TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
TITLE OF INVENTION: PORTIONS AND PRODUCTS
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,741
FILING DATE: 16-SEP-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Frommer Esq., William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454312-2460
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 864 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: amino acid
US-08-714-741-40

Query Match 17.1%; Score 7; DB 4; Length 864;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 KKEELTS 29
Db 296 KKEELTS 302

RESULT 10

US-08-714-741-41
Sequence 41, Application US/08714741
Patent No. 6500613

GENERAL INFORMATION:
APPLICANT: Briles, David E.
APPLICANT: McDaniel, Larry S.
APPLICANT: Swatlo, Edwin
APPLICANT: Yocher, Janet
APPLICANT: Crain, Marilyn J.
APPLICANT: Hollingshead, Susan
APPLICANT: Tart, Rebecca
APPLICANT: Brooke-Walter, Alexis
TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
TITLE OF INVENTION: PORTIONS AND PRODUCTS
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,741
FILING DATE: 16-SEP-1996
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Prommer Esq., William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454312-2460
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 1231 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: amino acid
US-08-714-741-41

Query Match 17.1%; Score 7; DB 4; Length 1231;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 KKEELTS 29
Db 125 KKEELTS 131

RESULT 11

US-09-513-999C-4343
Sequence 4343, Application US/09513999C
Patent No. 6783861
GENERAL INFORMATION:

APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclercq, A.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
Patent No. 6783861

FILE REFERENCE: 59, US2, REG
CURRENT APPLICATION NUMBER: US/09/513,999C
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 4343
LENGTH: 51
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SIGNAL
LOCATION: -36...-1
OTHER INFORMATION: score 5.4
OTHER INFORMATION: seq LDCYFGMSAVQSL/QL
US-09-513-999C-4343

Query Match 14.6%; Score 6; DB 4; Length 51;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 QSQULTA 36
Db 35 QSQULTA 40

RESULT 12

US-09-270-767-57083
Sequence 57083, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 57083
LENGTH: 56
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-270-767-57083

Query Match 14.6%; Score 6; DB 4; Length 56;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 SHQSOL 34
Db 10 SHQSOL 15

RESULT 13

US-09-621-976-5794
Sequence 5794, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm

Search completed: February 1, 2005, 15:35:43
 Job time : 18.5556 secs

; SEQ ID NO 5794
 ; LENGTH: 97
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SIGNAL
 ; LOCATION: -53...-1
 US-09-621-976-5794

Query Match 14.6%; Score 6; DB 4; Length 97;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 OSQULTA 36
 |||||
 Db 76 OSQULTA 81

RESULT 14
 US-09-270-767-41841
 ; Sequence 41841, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1998-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 41841
 ; LENGTH: 142
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 ; FEATURE:
 ; OTHER INFORMATION: Xaa means any amino acid
 US-09-270-767-41841

Query Match 14.6%; Score 6; DB 4; Length 142;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 29 SHOSQL 34
 |||||
 Db 10 SHOSQL 15

RESULT 15
 US-09-198-452A-6
 ; Sequence 6, Application US/09198452A
 ; Patent No. 6559294
 ; GENERAL INFORMATION:
 ; APPLICANT: Griffiths, R.
 ; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
 ; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
 ; FILE REFERENCE: 9710-003-999
 ; CURRENT APPLICATION NUMBER: US/09/198,452A
 ; CURRENT FILING DATE: 1998-11-24
 ; NUMBER OF SEQ ID NOS: 6849
 ; SEQ ID NO 6
 ; LENGTH: 143
 ; TYPE: PRT
 ; ORGANISM: Chlamydia pneumoniae
 US-09-198-452A-6

Query Match 14.6%; Score 6; DB 4; Length 143;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 SILOEH 18
 |||||
 Db 75 SILOEH 80

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 15:09:39 ; Search time 59.7778 Seconds
(without alignments)
247.799 Million cell updates/sec

Seq5ALA1854

Title: SE05ALA1854
Perfect score: 41
Sequence: 1 EKMKNMKMPMSILOEHIG.....MKKEELSHSQSLAFLEA 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1608061 seqs, 361289366 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1311380

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

1: Published Applications_AA:*
2: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	20	48.8	1149 17 US-10-128-558-167	Sequence 167, App
2	8	19.5	443 15 US-10-282-122A-70087	Sequence 70087, A
3	8	19.5	467 15 US-10-282-122A-68101	Sequence 68101, A
4	7	17.1	336 15 US-10-289-636-103	Sequence 103, App
5	7	17.1	431 14 US-10-254-995-3	Sequence 3, Appli
6	7	17.1	496 9 US-09-748-875-4	Sequence 4, Appli
7	7	17.1	496 10 US-09-298-523B-4	Sequence 4, Appli
8	7	17.1	589 9 US-09-748-875-14	Sequence 14, Appl
9	7	17.1	589 10 US-09-288-523B-14	Sequence 14, Appl
10	7	17.1	589 15 US-10-289-636-97	Sequence 97, Appl
11	7	17.1	643 15 US-10-289-636-95	Sequence 97, Appl
12	7	17.1	670 9 US-09-748-875-63	Sequence 63, Appl
13	7	17.1	670 10 US-09-298-523B-63	Sequence 63, Appl

14	7	17.1	690 9 US-09-748-875-61	Sequence 61, Appl
15	7	17.1	690 10 US-09-298-523B-61	Sequence 61, Appl
16	7	17.1	691 9 US-09-748-875-1	Sequence 1, Appli
17	7	17.1	691 10 US-09-288-523B-1	Sequence 1, Appli
18	7	17.1	701 9 US-09-748-875-62	Sequence 62, Appl
19	7	17.1	701 10 US-09-298-523B-62	Sequence 62, Appl
20	7	17.1	707 9 US-09-748-875-2	Sequence 2, Appli
21	7	17.1	707 10 US-09-298-523B-2	Sequence 2, Appli
22	7	17.1	711 9 US-09-748-875-3	Sequence 3, Appli
23	7	17.1	711 10 US-09-288-523B-3	Sequence 3, Appli
24	7	17.1	714 14 US-10-291-265-459	Sequence 459, App
25	7	17.1	929 9 US-09-748-875-60	Sequence 60, Appl
26	7	17.1	929 10 US-09-298-523B-60	Sequence 60, Appl
27	7	17.1	929 15 US-10-289-636-94	Sequence 94, Appl
28	6	14.6	54 16 US-10-437-863-136574	Sequence 136574, Ap
29	6	14.6	56 9 US-09-764-869-1049	Sequence 1049, Ap
30	6	14.6	56 10 US-09-764-891-5371	Sequence 5371, Ap
31	6	14.6	56 14 US-10-091-504-1049	Sequence 1049, Ap
32	6	14.6	56 15 US-10-227-577-1049	Sequence 1049, Ap
33	6	14.6	58 17 US-10-425-115-216216	Sequence 216216, Ap
34	6	14.6	58 17 US-10-425-115-356680	Sequence 356680, Ap
35	6	14.6	61 17 US-10-425-115-299684	Sequence 299684, Ap
36	6	14.6	61 17 US-10-425-115-299684	Sequence 299684, Ap
37	6	14.6	73 17 US-10-425-115-191450	Sequence 191450, A
38	6	14.6	84 14 US-10-029-386-29893	Sequence 29893, A
39	6	14.6	90 17 US-10-425-115-261981	Sequence 261981, A
40	6	14.6	92 17 US-10-425-115-189548	Sequence 189548, Ap
41	6	14.6	95 15 US-10-424-599-257004	Sequence 257004, Ap
42	6	14.6	95 15 US-10-276-774-1710	Sequence 1710, Ap
43	6	14.6	99 15 US-10-424-599-162535	Sequence 162535, Ap
44	6	14.6	99 15 US-10-424-599-204559	Sequence 204559, Ap
45	6	14.6	106 16 US-10-767-701-61193	Sequence 61193, A

ALIGNMENTS

RESULT 1
US-10-128-558-167
; Sequence 167, Application US/10128558
; Publication No. US20040219521A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Gezhi
; APPLICANT: Boyle, Bryan J
; APPLICANT: Dimanac, Radoje T
; TITLE OF INVENTION: Novel Nucleic Acids and
; FILE REFERENCE: 812A
; CURRENT APPLICATION NUMBER: US/10/128, 558
; CURRENT FILING DATE: 2002-04-22
; PRIOR APPLICATION NUMBER: US 60/339,453
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; Remaining Prior Application data removed - See File Wrapper or PALM.

```
; NUMBER OF SEQ ID NOS: 412
; SOFTWARE: pt_fl_genes Version 6.0
; SEQ ID NO 167
; LENGTH: 1149
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-128-558-167

Query Match          48.8%; Score 20; DB 17; Length 1149;
Best Local Similarity 100.0%; Pred. No. 8,2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      22 MKKEELTSHQSOLTAFFLEA 41
      |||||
Db      860 MKKEELTSHQSOLTAFFLEA 879

RESULT 2
US-10-282-122A-70087
; Sequence 70087, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70087
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Pseudomonas syringae
US-10-282-122A-70087

Query Match          19.5%; Score 8; DB 15; Length 443;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      34 LTAFFLEA 41
      |||||
```

```
Db      67 LTAFFLEA 74

RESULT 3
US-10-282-122A-68101
; Sequence 68101, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 68101
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Pseudomonas putida
US-10-282-122A-68101

Query Match          19.5%; Score 8; DB 15; Length 467;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      34 LTAFFLEA 41
      |||||
Db      92 LTAFFLEA 99

RESULT 4
US-10-299-636-103
; Sequence 103, Application US/10299636
; Publication No. US2004007847A1
; GENERAL INFORMATION:
; APPLICANT: Briles, David E
; APPLICANT: McDaniel, Larry S
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yother, Janet
; APPLICANT: Crain, Marilyn J
; APPLICANT: Hollingshead, Susan
```

APPLICANT: Tart, Rebecca
APPLICANT: Brooks-Walter, Alexis
TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEINS AND USES THEREOF
FILE REFERENCE: 57909/361
CURRENT APPLICATION NUMBER: US/10/299,636
CURRENT FILING DATE: 2002-11-19
PRIOR APPLICATION NUMBER: 08/714,741
PRIOR FILING DATE: 1996-09-16
PRIOR APPLICATION NUMBER: 08/529,055
PRIOR FILING DATE: 1995-09-15
NUMBER OF SEQ ID NOS: 111
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 103
LENGTH: 336
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-10-299-636-103

Query Match 17.1%; Score 7; DB 15; Length 336;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 KKEELTS 29
Db 56 KKEELTS 62

RESULT 5
US-10-254-995-3
Sequence 3, Application US/10254995
Publication No. US20030138447A1
GENERAL INFORMATION:
APPLICANT: Wizemann, Theresa M.
APPLICANT: Koenig, Scott
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/10/254,995
CURRENT FILING DATE: 2002-09-25
PRIOR APPLICATION NUMBER: US/09/286,981
PRIOR FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 3
LENGTH: 431
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-10-254-995-3

Query Match 17.1%; Score 7; DB 14; Length 431;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 KKEELTS 29
Db 74 KKEELTS 80

RESULT 6
US-09-748-875-4
Sequence 4, Application US/09748875
Publication No. US20010016200A1
GENERAL INFORMATION:
APPLICANT: BRILES et al.
TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
FILE REFERENCE: 454312-3140
CURRENT APPLICATION NUMBER: US/09/748,875
CURRENT FILING DATE: 2000-12-26
PRIOR APPLICATION NUMBER: 09/298,523
PRIOR FILING DATE: 1999-04-23

NUMBER OF SEQ ID NOS: 78
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 4
LENGTH: 496
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-748-875-4

Query Match 17.1%; Score 7; DB 9; Length 496;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 KKEELTS 29
Db 120 KKEELTS 126

RESULT 7
US-09-298-523B-4
Sequence 4, Application US/09298523B
Publication No. US20030059438A1
GENERAL INFORMATION:
APPLICANT: BRILES et al.
TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
FILE REFERENCE: 454312-3140
CURRENT APPLICATION NUMBER: US/09/298,523B
CURRENT FILING DATE: 1999-04-23
NUMBER OF SEQ ID NOS: 78
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 4
LENGTH: 496
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-298-523B-4

Query Match 17.1%; Score 7; DB 10; Length 496;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 KKEELTS 29
Db 120 KKEELTS 126

RESULT 8
US-09-748-875-14
Sequence 14, Application US/09748875
Publication No. US20010016200A1
GENERAL INFORMATION:
APPLICANT: BRILES et al.
TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
FILE REFERENCE: 454312-3140
CURRENT APPLICATION NUMBER: US/09/748,875
CURRENT FILING DATE: 2000-12-26
PRIOR APPLICATION NUMBER: 09/298,523
PRIOR FILING DATE: 1999-04-23
NUMBER OF SEQ ID NOS: 78
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 14
LENGTH: 589
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-748-875-14

Query Match 17.1%; Score 7; DB 9; Length 589;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 KKEELTS 29
Db 324 KKEELTS 330

```
RESULT 9
US-09-298-523B-14
; Sequence 14, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; PRIOR FILING DATE: 1999-04-23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 589
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-14
```

```
Query Match 17.1%; Score 7; DB 10; Length 589;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 23 KKEELTS 29
Db 324 KKEELTS 330
```

```
RESULT 10
US-10-299-636-97
; Sequence 97, Application US/10299636
; Publication No. US20040077847A1
; GENERAL INFORMATION:
; APPLICANT: Briles, David E
; APPLICANT: McDaniel, Larry S
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yoether, Janet
; APPLICANT: Crain, Marilyn J
; APPLICANT: Hollingshead, Susan
; APPLICANT: Tart, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEINS AND USES THEREOF
; FILE REFERENCE: 57909/361
; CURRENT APPLICATION NUMBER: US/10/299,636
; PRIOR FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: 08/714,741
; PRIOR FILING DATE: 1996-09-16
; PRIOR APPLICATION NUMBER: 08/529,055
; PRIOR FILING DATE: 1995-09-15
; NUMBER OF SEQ ID NOS: 111
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 97
; LENGTH: 589
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-299-636-97
```

```
Query Match 17.1%; Score 7; DB 15; Length 589;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 23 KKEELTS 29
Db 324 KKEELTS 330
```

```
RESULT 11
US-10-299-636-95
; Sequence 95, Application US/10299636
; Publication No. US20040077847A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Briles, David E
; APPLICANT: McDaniel, Larry S
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yoether, Janet
; APPLICANT: Crain, Marilyn J
; APPLICANT: Hollingshead, Susan
; APPLICANT: Tart, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEINS AND USES THEREOF
; FILE REFERENCE: 57909/361
; CURRENT APPLICATION NUMBER: US/10/299,636
; PRIOR FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: 08/714,741
; PRIOR FILING DATE: 1996-09-16
; PRIOR APPLICATION NUMBER: 08/529,055
; PRIOR FILING DATE: 1995-09-15
; NUMBER OF SEQ ID NOS: 111
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 95
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-299-636-95
```

```
Query Match 17.1%; Score 7; DB 15; Length 643;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 23 KKEELTS 29
Db 76 KKEELTS 82
```

```
RESULT 12
US-09-748-875-63
; Sequence 63, Application US/09748875
; Publication No. US20010016200A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/748,875
; PRIOR FILING DATE: 2000-12-26
; PRIOR APPLICATION NUMBER: 09/298,523
; PRIOR FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 670
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-748-875-63
```

```
Query Match 17.1%; Score 7; DB 9; Length 670;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 23 KKEELTS 29
Db 125 KKEELTS 131
```

```
RESULT 13
US-09-298-523B-63
; Sequence 63, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
```

```

; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 63
; LENGTH: 670
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-63

```

```

Query Match      17.1%; Score 7; DB 10; Length 670;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      23 KKEELTS 29
        |||||
Db      125 KKEELTS 131

```

```

RESULT 14
US-09-748-875-61
; Sequence 61, Application US/09748875
; Publication No. US20010016200A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/748,875
; CURRENT FILING DATE: 2000-12-26
; PRIOR APPLICATION NUMBER: 09/298,523
; PRIOR FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 61
; LENGTH: 690
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-748-875-61

```

```

Query Match      17.1%; Score 7; DB 9; Length 690;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      23 KKEELTS 29
        |||||
Db      361 KKEELTS 367

```

```

RESULT 15
US-09-298-523B-61
; Sequence 61, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 61
; LENGTH: 690
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-61

```

```

Query Match      17.1%; Score 7; DB 10; Length 690;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      23 KKEELTS 29
        |||||

```

Db 361 KKEELTS 367

Search completed: February 1, 2005, 15:44:48
Job time : 60.7778 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:55:01 ; Search time 13.5556 Seconds
(without alignment)
291.016 Million cell updates/sec

Title: SEQ5ALA1854

Perfect score: 41

Sequence: 1 EKWMKNMGPFMSIAGEHIG.....MKKEELTSHQSQTAFLEA 41

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size: 0

Total number of hits satisfying chosen parameters: 279530

Minimum DB seq length: 21

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	19.5	467	2	AC0546
2	8	19.5	488	2	T47273
3	7	17.1	276	2	T08859
4	7	17.1	417	2	D84556
5	7	17.1	537	2	A38170
6	7	17.1	960	2	544812
7	7	17.1	1196	2	H86389
8	7	17.1	1478	2	S78131
9	6	14.6	40	2	T06840
10	6	14.6	69	2	S04666
11	6	14.6	75	2	S75054
12	6	14.6	107	1	A26622
13	6	14.6	134	2	S75307
14	6	14.6	137	2	T46439
15	6	14.6	143	2	B86624
16	6	14.6	143	2	G72000
17	6	14.6	160	2	T44921
18	6	14.6	189	2	H86520
19	6	14.6	219	1	Z0Z0MR
20	6	14.6	223	2	B82911
21	6	14.6	228	2	T15530
22	6	14.6	228	2	AC3204
23	6	14.6	230	2	T03046
24	6	14.6	237	2	G70066
25	6	14.6	241	2	G97416
26	6	14.6	241	2	A12634
27	6	14.6	274	1	F40362
28	6	14.6	276	2	S27641
29	6	14.6	280	2	D86193

30	6	14.6	290	2	AB1176	fructokinases homo
31	6	14.6	290	2	A11533	fructokinases homo
32	6	14.6	319	2	H69882	deacetylase homo
33	6	14.6	351	2	T03153	hypothetical prote
34	6	14.6	358	2	J00596	nucleosid prote
35	6	14.6	372	2	S76427	hypothetical prote
36	6	14.6	377	2	AD1937	permease protein o
37	6	14.6	397	2	G83699	malate oxidoreduct
38	6	14.6	398	2	S13269	translacion initia
39	6	14.6	402	2	S30278	translacion initia
40	6	14.6	406	1	F1MS4A	translacion initia
41	6	14.6	406	2	S33681	translacion initia
42	6	14.6	407	2	S00985	translacion initia
43	6	14.6	428	2	B81531	conserved hypothet
44	6	14.6	432	2	D72008	CT850 hypothetical
45	6	14.6	432	2	E86616	CT850 hypothetical

ALIGNMENTS

RESULT 1
AG0546
probable terminal oxidase chain I [imported] - Salmonella enterica subsp. enterica serov
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AG0546
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AG0546
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-467 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08815.1; PID:G16501629; GSPDB:GN00176
A:Gene: STY0392
C:Superfamily: cytochrome d complex terminal oxidase chain I

Query Match 19.5% Score 8; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 LTAFLEA 41
DB 102 LTAFLEA 109

RESULT 2
T47273
cyanide insensitive terminal oxidase chain cioa [imported] - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C:Accession: T47273; G83155
R:Cunningham, L.; Pitt, M.; Williams, H.D.
Mol. Microbiol. 24, 579-591, 1997
A:Title: The cioAB genes from Pseudomonas aeruginosa code for a novel cyanide-insensitiv
A:Reference number: Z24440; MUID:9732403; PMID:9179851
A:Accession: T47273
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-488 <CUN>
A:Cross-references: UNIPROT:O07440; EMBL:Y10528; NID:G2208963; PIDN:CAA71555.1; PID:G220
A:Experimental source: Estrain PA01, substrain PA06049
R:Stover, C.K.; Pham, X.O.; Ervin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Lapid, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000

A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
A:Reference number: A82950; UID:20437337; PMID:10984043
A:Accession: G83155
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-488 <STO>
A:Cross-references: GB:AE004810; GB:AE004091; NID:g9950106; PIDN:AAG07317.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: cloA; PA3930
C:Superfamily: cytochrome d complex terminal oxidase chain I

Query Match 19.5%; Score 8; DB 2; Length 488;
Best Local Similarity 100.0%; Pred. No. 0.82;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 LTAFFLEA 41
DB 103 LTAFFLEA 110

RESULT 3
T08859
hypothetical protein A_TMO17A05.4 - *Arabidopsis thaliana*
C:Species: *Arabidopsis thaliana* (mouse-ear cress)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 23-Jul-1999
C:Accession: T08859
R:Waterston, R.
submitted to the EMBL Data Library, October 1997
A:Reference number: Z16500
A:Accession: T08859
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-276 <WAT>
A:Cross-references: EMBL:AF024504; NID:g2435510; PID:g2435512
C:Genetics:
A:Map position: 4
A:Introns: 137/3
A:Note: A_TMO17A05.4

Query Match 17.1%; Score 7; DB 2; Length 276;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 KKEELTS 29
DB 243 KKEELTS 249

RESULT 4
D84556
hypothetical protein At2g17780 [imported] - *Arabidopsis thaliana*
C:Species: *Arabidopsis thaliana* (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: D84556
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umayam, L.; Tallon, L.;
Euse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
A:Reference number: A84420; UID:20083487; PMID:10617197
A:Accession: D84556
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-417 <STO>
A:Cross-references: GB:AE002093; NID:g6598810; PIDN:AAB80787.2; GSPDB:GN00139
C:Genetics:
A:Gene: At2g17780
A:Map position: 2

Query Match 17.1%; Score 7; DB 2; Length 417;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 KKEELTS 29
DB 243 KKEELTS 249

RESULT 5
A38170
cytochrome d complex terminal oxidase (EC 1.10.3.-) chain I - *Azotobacter vinelandii*
C:Species: *Azotobacter vinelandii*
C:Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 09-Jul-2004
C:Accession: A38170; A41578
R:Moshiri, F.; Chawla, A.; Maier, R.J.
J. Bacteriol. 173, 6230-6241, 1991
A:Title: Cloning, characterization, and expression in *Escherichia coli* of the genes encod
A:Reference number: A38170; UID:92011387; PMID:1655703
A:Accession: A38170
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-537 <MOS>
A:Cross-references: UNIPROT:Q09049; GB:S57066; NID:g236633; PIDN:AAB19986.1; PID:g236634
R:Moshiri, F.; Smith, E.G.; Taormino, V.P.; Maier, R.J.
J. Biol. Chem. 266, 23169-23174, 1991
A:Title: Transcriptional regulation of cytochrome d in nitrogen-fixing *Azotobacter vinel*
A:Reference number: A41578; UID:92078187; PMID:1660468
A:Accession: A41578
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-13 <MO2>
A:Cross-references: GB:M77787
A:Note: the authors translated the codon CAG for residue 13 as G1y
C:Superfamily: cytochrome d complex terminal oxidase chain I
C:Keywords: oxidoreductase

Query Match 17.1%; Score 7; DB 2; Length 537;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 LTAFFLE 40
DB 103 LTAFFLE 109

RESULT 6
S44812
F44B9.6 protein - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 04-Mar-2000
C:Accession: S44812
R:Antonaci-Fulton, L.
submitted to the EMBL Data Library, September 1993
A:Description: Sequence of the C. elegans cosmid F44B9.
A:Reference number: S44807
A:Accession: S44812
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-960 <ANT>
A:Cross-references: EMBL:L23648; NID:g388505; PID:g388591
C:Genetics:
A:Introns: 33/3; 66/3; 261/2; 299/3; 479/2; 615/3; 750/3; 903/2; 925/3
C:Superfamily: *Caenorhabditis elegans* F44B9.6 protein

Query Match 17.1%; Score 7; DB 2; Length 960;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 ELTSHOS 32
DB 478 ELTSHOS 484

RESULT 7
H86389

hypothetical protein F28823.2 - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
 C/Accession: H66389
 R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiet, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 Auer, N.F.; Hughes, B.; Hultzer, L.
 Nature 408, 816-820, 2000
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, J.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marzilli,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A/Reference number: A66141; MUID:21016719; PMID:11130712
 A/Accession: H66389
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-1196 <STO>
 A/Cross-references: UNIPROT:Q9C678; GB:AE005172; NID:g11079511; PIDN:AA629222.1; GSPDB:C
 C/Genetics:
 A/Map position: 1

Query Match 17.1%; Score 7; DB 2; Length 1196;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 22 MKKEELT 28
 |||||
 Db 349 MKKEELT 355

RESULT 8

DNA-directed RNA polymerase (EC 2.7.7.6) chain beta - Reclinomonas americana (ATCC 50394
 C/Species: mitochondrion Reclinomonas americana
 A/Variety: ATCC 50394
 C/Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 09-Jul-2004
 C/Accession: S78131
 R/Liang, B.F.; Burger, G.; O'Kelly, C.J.; Cedergren, R.; Golding, G.B.; Lemieux, C.; Sank
 Nature 387, 493-497, 1997
 A/Title: An ancestral mitochondrial DNA resembling a eubacterial genome in miniature.
 A/Reference number: S78127; MUID:97311393; PMID:9168110
 A/Accession: S78131
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-1478 <LAN>
 A/Cross-references: UNIPROT:O21237; EMBL:AF007261; NID:g2258325; PIDN:AA011864.1; PID:g2
 A/Experimental source: ATCC 50394
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1997
 C/Genetics:
 A/Gene: rpoB
 A/Genome: mitochondrion
 C/Superfamily: DNA-directed RNA polymerase beta chain
 C/Keywords: mitochondrion; nucleotidyltransferase; transcription

Query Match 17.1%; Score 7; DB 2; Length 1478;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 23 KKEELTS 29
 |||||
 Db 426 KKEELTS 432

RESULT 9

photosystem I chain IX - Cyanophora paradoxa cyanelle
 T06840
 C/Species: cyanelle Cyanophora paradoxa
 C/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004
 C/Accession: T06840
 R/Streval, V.L.; Michalowski, C.B.; Luffelhardt, W.; Bohner, H.J.; Bryant, D.A.
 submitted to the EMBL Data Library, July 1995

A/Description: Nucleotide sequence of the cyanelle genome from Cyanophora paradoxa.
 A/Reference number: Z15840
 A/Accession: T06840
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-70 <STI>
 A/Cross-references: UNIPROT:P48117; EMBL:U30821; NID:g1016083; PIDN:AAA81183.1; PID:g101
 A/Experimental source: strain Pringsheim LB555
 C/Genetics:
 A/Gene: psal
 A/Genome: cyanelle
 C/Superfamily: photosystem I protein psal
 C/Keywords: cyanelle; photosynthesis; photosystem I; thylakoid

Query Match 14.6%; Score 6; DB 2; Length 40;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 34 LTAFFL 39
 |||||
 Db 16 LTAFFL 21

RESULT 10

S04666
 hypothetical protein 1 - Rhodospseudomonas blastica (fragment)
 C/Species: Rhodospseudomonas blastica
 C/Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
 C/Accession: S04666
 R/Tydulewicz, V.L.J.; Falk, G.; Walker, J.E.
 J. Mol. Biol. 179, 185-214, 1984
 A/Title: Rhodospseudomonas blastica atp operon. Nucleotide sequence and transcription.
 A/Reference number: S04666; MUID:85058188; PMID:6209404
 A/Accession: S04666
 A/Status: not compared with conceptual translation
 A/Molecule type: DNA
 A/Residues: 1-69 <TVB>
 A/Cross-references: UNIPROT:P05443

Query Match 14.6%; Score 6; DB 2; Length 69;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 25 EELTSH 30
 |||||
 Db 3 EELTSH 8

RESULT 11

S75054
 hypothetical protein sbl127 - Synechocystis sp. (strain PCC 6803)
 C/Species: Synechocystis sp.
 A/Variety: PCC 6803
 C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
 C/Accession: S75054
 R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O.K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yaeuda
 DNA Res. 3, 109-116, 1996
 A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
 sp.
 A/Reference number: S74322; MUID:97061201; PMID:8905231
 A/Accession: S75054
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-75 <KAN>
 A/Cross-references: UNIPROT:P73857; EMBL:D90910; GB:AB001339; NID:g1652956; PIDN:BA41791
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C/Genetics:
 A/Start codon: GTG

Query Match 14.6%; Score 6; DB 2; Length 75;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 KKEBLT 28
 |||||
 Db 67 KKEBLT 72

RESULT 12

A26622
 Chloredoxin - Chromatium vinosum
 C/Species: Chromatium vinosum
 C/Date: 31-Mar-1988 #sequence_revision 26-May-1994 #text_change 16-Aug-2004
 C/Accession: A26622
 R/Johnson, R.S.; Blemann, K.
 Biochemistry 26, 1209-1214, 1987
 A/Title: The primary structure of thioedoxin from Chromatium vinosum determined by high
 A/Reference number: A26622; MUID:87185419; PMID:3567166
 A/Accession: A26622
 A/Molecule type: protein
 A/Residues: 1-107 <JOH>
 A/Cross-references: UNIPROT:P09857
 A/Note: unidentified residues are Ile or Leu
 C/Species: Thioedoxin; Chloredoxin homology
 C/Keywords: heat-stable protein; redox-active disulfide
 F:10-93/Domain: thioedoxin homology <THR>
 F:32-35/Disulfide bonds: redox-active #status predicted

Query Match 14.6%; Score 6; DB 1; Length 107;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 SOLTAF 37
 |||||
 Db 97 SOLTAF 102

RESULT 13

S75307
 ferredoxin [2Fe-2S] s111584 [similarity] - Synechocystis sp. (strain PCC 6803)
 C/Species: Synechocystis sp.
 A/Variety: PCC 6803
 C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
 C/Accession: S75307
 R/Kaneke, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O. K.; Okumura, S.; Shimpo, S.; Takuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
 sp.
 A/Reference number: S74322; MUID:97061201; PMID:8905231
 A/Accession: S75307
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-134 <KAN>
 A/Cross-references: UNIPROT:P71395; EMBL:D90904; GB:AB001339; NID:g1652225; PIDN:BA11722
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C/Species: ferredoxin [2Fe-2S], Clostridium type
 C/Keywords: 2Fe-2S; electron transfer; iron-sulfur protein; metalloprotein
 F:32,40,80,84/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match 14.6%; Score 6; DB 2; Length 134;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 ILQEH 19
 |||||
 Db 110 ILQEH 115

RESULT 14

T46439
 hypothetical protein DKFZp434M0326.1 - human
 C/Species: Homo sapiens (man)
 C/Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
 C/Accession: T46439

R/Ansorge, W.; Winkner, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, January 2000
 A/Reference number: Z23028
 A/Accession: T46439

A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-137 <AAA>
 A/Cross-references: UNIPROT:Q9NSX2; EMBL:AL137681
 A/Experimental source: adult testis; clone DKFZp434M0326
 C/Genetics:
 A/Note: DKFZp434M0326.1

Query Match 14.6%; Score 6; DB 2; Length 137;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 KKEBLT 28
 |||||
 Db 22 KKEBLT 27

RESULT 15

B86624
 HTT transcription regulator [imported] - Chlamydomonas reinhardtii (strain J138)
 C/Species: Chlamydomonas reinhardtii
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
 C/Accession: B86624
 R/Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Iet
 Nucleic Acids Res. 28, 2311-2314, 2000
 A/Title: Comparison of whole genome sequences of Chlamydomonas reinhardtii J138
 A/Reference number: A86491; MUID:20330349; PMID:10871362
 A/Accession: B86624
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-143 <STO>
 A/Cross-references: UNIPROT:Q9Z6J0; GB:BA000008; NID:g8979442; PIDN:BA99276.1; GSPDB:GNV
 A/Experimental source: strain J138
 C/Genetics:
 A/Gene: yfgA

Query Match 14.6%; Score 6; DB 2; Length 143;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 SILOEH 18
 |||||
 Db 75 SILOEH 80

Search completed: February 1, 2005, 15:33:02
 Job time: 15.5556 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:53:01 ; Search time 69.7778 Seconds
(without alignments)
210.782 Million cell updates/sec

Title: SEQ5ASNL694

Perfect score: 41
Sequence: 1 LKLLCKNFGAENPDFFVPL.....TAVKLIAPERKEKNVLGSA 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2002273 seqs, 358729239 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1316031

Minimum DB seq length: 21

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	1149	7	ADE08012
2	20	48.8	515	2	AAW54099
3	20	48.8	2144	4	AAH85029
4	7	17.1	118	4	AAH87229
5	7	17.1	308	6	ABU24428
6	7	17.1	732	4	ABG17089
7	7	17.1	976	8	ADN73151
8	7	17.1	2000	8	ADN04625
9	7	17.1	4618	4	AAH39043
10	7	14.6	32	7	ADL46346
11	6	14.6	32	4	AAH89273
12	6	14.6	47	4	AAH20415
13	6	14.6	47	4	ABH41137
14	6	14.6	47	4	AAH34913
15	6	14.6	47	4	ABH25176
16	6	14.6	47	4	AAH74797
17	6	14.6	47	4	AAH61993
18	6	14.6	47	4	ABG56579
19	6	14.6	47	5	ABG44588
20	6	14.6	50	4	ABH03116
21	6	14.6	50	6	ABU12410
22	6	14.6	50	8	ADJ28436
23	6	14.6	51	3	AAH01477
24	6	14.6	61	5	ABJ10298
25	6	14.6	61	8	ADK47657

26	6	14.6	62	6	ABU20127	Abu20127 Protein e
27	6	14.6	66	3	AAH19461	AaH19461 Arabidops
28	6	14.6	67	2	AAH32985	AaH32985 Encoded b
29	6	14.6	70	5	AAH25974	AaH25974 Human pro
30	6	14.6	70	7	ADH62559	AdH62559 Cardiac m
31	6	14.6	81	4	AAU32861	AaU32861 Novel hum
32	6	14.6	86	6	ABP79600	AbP79600 N. gonorr
33	6	14.6	87	3	AAH52529	AaH52529 Human 5'
34	6	14.6	87	3	AAH00535	AaH00535 Human sec
35	6	14.6	88	3	AAH14126	AbH14126 Bordetell
36	6	14.6	93	5	AAU91110	AaU91110 Human sec
37	6	14.6	95	4	AAH25936	AaH25936 Human pro
38	6	14.6	98	8	ADL05626	AdL05626 M. catarr
39	6	14.6	113	3	AAH57341	AaH57341 Arabidops
40	6	14.6	119	5	AAU91087	AaU91087 Human sec
41	6	14.6	119	5	ABH54604	AbH54604 Lactococc
42	6	14.6	121	5	ABH01787	AbH01787 Human bre
43	6	14.6	126	5	AAU91137	AaU91137 Human sec
44	6	14.6	128	3	AAH57954	AaH57954 Human tra
45	6	14.6	130	3	AAH41161	AbH41161 Human ORF

ALIGNMENTS

RESULT 1	
ADH08012	ADH08012 standard; protein; 1149 AA.
XX	
AC	ADH08012;
XX	
DT	29-JAN-2004 (first entry)
XX	
DE	Novel protein (useful for identifying genetic disorders) #167.
XX	
KW	novel gene; novel protein; tissue marker; molecular weight marker;
KW	chromosome marker; genetic disorder.
XX	
OS	Unidentified.
XX	
FN	WO2003054152-A2.
PD	03-JUL-2003.
XX	
PF	10-DEC-2002; 2002WO-US039555.
XX	
PR	10-DEC-2001; 2001US-0339739P.
PR	11-DEC-2001; 2001US-0339453P.
PR	14-MAR-2002; 2002US-0365091P.
PR	14-MAR-2002; 2002US-0365384P.
PR	12-APR-2002; 2002US-0372381P.
PR	12-APR-2002; 2002US-0372615P.
PR	22-APR-2002; 2002US-00128558.
PR	24-APR-2002; 2002US-0376045P.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
PI	Ghosh M, Xue AJ, Wehman T, Weng G, Zhou P, Drmanac RT, Wang Z;
PI	Ma Y, Wang D, Chen R, Xu C, Boyle BJ;
XX	
DR	WPI: 2003-569235/53.
XX	
DR	N-PSDB; ADH07101.
XX	
PT	New polynucleotides, useful for expressing recombinant proteins for
PT	analysis, characterization or therapeutic use, or as markers for tissues
PT	in which the corresponding protein is preferentially expressed.
XX	
PS	Claim 20; SEQ ID NO 1078; 1177bp; English.
XX	
CC	The invention comprises the amino acid and coding sequences of novel
CC	proteins. The DNA and protein sequences of the invention are useful as:
CC	markers for tissues in which the corresponding protein is preferentially

expressed; as molecular weight markers on gels; as chromosome markers or tags; to identify chromosomes or to map related gene positions; and to compare with endogenous DNA sequences in patients to identify potential genetic disorders. The present amino acid sequence represents a protein of the invention.

Sequence 1149 AA;

Query Match Best Local Similarity 100.0%; Score 41; DB 7; Length 1149;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 LKLLCKNFGAENPDFFVPLNTAVKLIAPERKEKNVLGSA 41
732 LKLLCKNFGAENPDFFVPLNTAVKLIAPERKEKNVLGSA 772

RESULT 2
AAM54099
ID AAM54099 standard; protein; 515 AA.

AAW54099;

28-SEP-1998 (first entry)

Homo sapiens BAP28 sequence.

BARD1; ring protein; BRCA1; breast cancer; risk; diagnosis.

Homo sapiens.

MO9812327-A2.

26-MAR-1998.

19-SEP-1997; 97MO-US016842.

20-SEP-1996; 96US-0025296P.

03-APR-1997; 97US-0042611P.

04-APR-1997; 97US-0042988P.

(TEXA) UNIV TEXAS SYSTEM.

Bowcock AM, Baer R;

WPI; 1998-230317/20.

N-PSDB; AAV24135.

DNA sequence encoding BARD1, B123, BE2, BE14, BE31 or BE445 - which as breast cancer antigen, BRCA1, binding proteins are useful to identify patient having or at risk of developing cancer.

Disclosure; Page 287-288; 348pp; English.

The sequence is that of a protein which can be used in the preparation of the recombinant breast cancer antigen, BRCA1, binding proteins BARD1, B123, BE2, BE14, BE31 or BE445, or a composition for the detection of a BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid sequence, specifically a wild type BARD1 composition for the detection or purification of BRCA1, useful to identify a patient having, or at risk of developing cancer. BARD1 can be used in the preparation of an anti-BARD1 antibody, and in the detection and purification of a BRCA1 protein. BARD1, B123, BE2, BE14, BE31 or BE445 can be used in the identification of a binding protein agonist or antagonist that alters the binding of BARD1, B123, BE2, BE14, BE31 or BE445 to BRCA1 or the biological activity of the BRCA1-BARD1, B123, BE2, BE14, BE31 or BE445 complex. The antibodies can be used to detect BARD1, B123, BE2, BE14, BE31 or BE445, a specific anti-BARD1 antibody can be used to identify a patient having or at risk of developing cancer.

Sequence 515 AA;

Query Match 48.8%; Score 20; DB 2; Length 515;

Best Local Similarity 100.0%; Pred. No. 1,3e-12; Mismatches 0; Indels 0; Gaps 0;

1 LKLLCKNFGAENPDFFVPL 20
45 LKLLCKNFGAENPDFFVPL 64

RESULT 3
AAB85029
ID AAB85029 standard; protein; 2144 AA.

AAB85029;

06-AUG-2001 (first entry)

Protein encoded by BAP28 cDNA consisting of exons 1 to 45.

BAP28; prostate; tumour; cancer; diagnostic; genetic analysis.

Homo sapiens.

Key Location/Qualifiers

FT Misc-difference 1694 /label= Ser or Asn

FT Misc-difference 1854 /label= Ala or Val

FT Misc-difference 1967 /label= Asp or Asn

FT Misc-difference 2017 /label= Gly or Glu

MO200100669-A2.

04-JAN-2001.

23-JUN-2000; 2000MO-IB001183.

25-JUN-1999; 99US-0141323P.

18-JAN-2000; 2000US-0176860P.

(GENSET) GENSET.

Barry C, Bougueleret L, Chumakov I, Cohen-Akenine A;

WPI; 2001-367032/38.

N-PSDB; AAF83909, AAF83910.

New BAP28 polynucleotides and polypeptides overexpressed in prostate cancer cells for diagnosing prostate tumors, e.g. by hybridization or polymerase chain reaction assays.

Claim 14; Page 297-304; 349pp; English.

The invention is directed to BAP28 polypeptides, BAP28 polynucleotide sequences and regulatory region located at the 3' and 5' ends of the BAP28 coding region. The BAP28 polypeptides can be expressed by standard recombinant methodology. BAP28 polynucleotides and polypeptides have been found to be over expressed in prostate tumour cells, therefore levels of BAP28 expression and/or activity may be assayed (e.g. by polymerase chain reaction (PCR)) to diagnose patient suffering from or susceptible to prostate cancer. Antibodies specific for the BAP28 polypeptides are useful as diagnostic reagents. Biallelic markers of the BAP28 gene are useful in genetic analysis. The present sequence represents a protein encoded by a first cDNA sequence of the BAP28 gene consisting of the exons 1 to 45

Sequence 2144 AA;

Query Match 48.8%; Score 20; DB 4; Length 2144;
Best Local Similarity 100.0%; Pred. No. 4.2e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LKLCNKGAEPPDFVPEVL 20
|||
Db 1674 LKLCNKGAEPPDFVPEVL 1693

RESULT 4
ID AAM87229 standard; protein; 118 AA.
XX AAM87229,
AC AAM87229,
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen SEQ ID NO:14822.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.
XX
OS Homo sapiens.
PN WO200157182-A2.
PD
XX
XX 09-AUG-2001.
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220966P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 23-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0228287P.
PR 01-SEP-2000; 2000US-0228343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.

PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235836P.
PR 27-SEP-2000; 2000US-0235837P.
PR 29-SEP-2000; 2000US-0236337P.
PR 29-SEP-2000; 2000US-0236357P.
PR 29-SEP-2000; 2000US-0236358P.
PR 29-SEP-2000; 2000US-0236359P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.

PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI; 2001-483426/52.
 DR N-PSDB; AAK60010.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and metastasis.
 XX
 PS Claim 11; SEQ ID NO 14822; 3071bp + Sequence Listing; English.
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (1)
 CC amino acid sequences given in AAK62170 to AAK61921. (1) have cytostratic
 CC activity, and can be used in gene therapy and vaccine production. (1)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (1) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC to affect the activity of (1) by expressing inactive proteins or to
 CC supplement the patient's own production of (1). Additionally, (1)
 CC polynucleotides may be used to produce the secreted (1), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (1) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK7654 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169
 CC represent sequences used in the exemplification of the present invention
 CC
 XX
 SQ Sequence 118 AA;
 Query Match 17.1%; Score 7; DB 4; Length 118;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 31 RKEKKNV 37
 |||||
 DB 52 RKEKKNV 58
 RESULT 5
 ABU24428
 ID ABU24428 standard; protein; 308 AA.
 XX
 AC ABU24428;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by prokaryotic essential gene #9955.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Clostridium botulinum.
 XX
 PN WO20027183-A2.

XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0382699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 DR WPI; 2003-029926/02.
 DR N-PSDB; ACA28298.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 52352; 1766bp; English.
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the antisense
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-regulated gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 XX
 SQ Sequence 308 AA;
 Query Match 17.1%; Score 7; DB 6; Length 308;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 35 KNTLGSA 41
 |||||
 DB 298 KNTLGSA 304
 RESULT 6
 ABG17089
 ID ABG17089 standard; protein; 732 AA.
 XX
 AC ABG17089;

XX 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #17080.
 XX
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Dermanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS81276.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 47448; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptides and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 732 AA;
 XX
 Query Match 17.1%; Score 7; DB 4; Length 732;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 15 PFVPLN 21
 Db 289 PFVPLN 295
 XX
 RESULT 7
 ADN73151
 ID ADN73151 standard; protein; 976 AA.
 XX
 AC ADN73151;
 XX
 DT 15-JUL-2004 (first entry)
 XX

XX Thale cress protein upregulated in E2Fa/Dpa expressing plants SeqID 1046.
 DE plant; transgenic; E2Fa/Dpa transcription factor; growth regulator;
 KW animal feed product; thale cress; cell wall biosynthesis;
 KW nitrogen metabolism; carbon metabolism.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO2004035798-A2.
 XX
 PD 29-APR-2004.
 XX
 PF 20-OCT-2003; 2003WO-EP011658.
 XX
 PR 18-OCT-2002; 2002EP-00079408.
 XX
 PA (CROP-) CROPDESIGN NV.
 XX
 PI Inze D, De Veylder L, Vlieghe K;
 XX
 DR WPI; 2004-348466/32.
 DR N-PSDB; ADN73150.
 XX
 PT Altering plant characteristics, useful for producing plants for enzyme or
 PT pharmaceutical production comprises modifying in a plant, expression of
 PT one or more nucleic acids and/or modifying level or activity of one or
 PT more proteins.
 XX
 PS Claim 1; SEQ ID NO 1046; 134pp; English.
 XX
 CC This invention relates to a novel method for altering one or more plant
 CC characteristics. Specifically, it refers to identifying genes that are up
 CC - or down-regulated in transgenic plants overexpressing the heterodimeric
 CC E2Fa/Dpa transcription factor of Arabidopsis and using these sequences to
 CC alter plant characteristics accordingly. The present invention describes
 CC generating transgenic plants for the production of growth regulators,
 CC enzymes, therapeutics, pharmaceuticals and animal feed products, where
 CC the altered plant characteristics are selected from increased yield or
 CC biomass, enhanced survival capacity, stress tolerance, plant architecture
 CC or physiology, altered endoreplication, biochemistry, signal
 CC transduction, storage lipid mobilisation and/or altered photosynthesis,
 CC each relative to the corresponding wild type plants. Accordingly, these
 CC sequences can also be useful as positive or negative selectable markers
 CC during transformation of cells or tissues. The identified genes play a
 CC role in a variety of biological processes such as DNA replication, cell
 CC wall biosynthesis, nitrogen and/or carbon metabolism or they function as
 CC transcription factors. This polypeptide sequence is thale cress protein
 CC expressed by a gene upregulated 1.3 fold or more in plants overexpressing
 CC the E2Fa/Dpa transcription factor, given in an exemplification of the
 CC invention.
 XX
 SQ Sequence 976 AA;
 XX
 Query Match 17.1%; Score 7; DB 8; Length 976;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 32 KEKNVL 38
 Db 474 KEKNVL 480
 XX
 RESULT 8
 ADN04625
 ID ADN04625 standard; protein; 2000 AA.
 XX
 AC ADN04625;
 XX
 DT 01-JUL-2004 (first entry)
 DE Antisporadic protein sequence #498.
 XX

KW antipsoriatic; gene therapy; psoriasis; diagnosis.
 XX
 OS Homo sapiens.
 XX
 PN WO2004028479-A2.
 XX
 PD 08-APR-2004.
 XX
 PF 25-SEP-2003; 2003WO-US030907.
 XX
 PR 25-SEP-2002; 2002US-0414006P.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
 PI Mu TD;
 XX
 DR WPI; 2004-305105/28.
 DR N-PSDB; ADN04624.
 XX
 PT New PRO nucleic acid or polypeptide, useful for preparing a
 PT pharmaceutical composition for diagnosing or treating psoriasis in a
 PT mammal.
 XX
 PS Claim 9, SEQ ID NO 1019, 3069pp, English.
 XX
 CC The invention relates to novel polynucleotide and polypeptides for
 CC treating psoriasis or a sequence having at least 80% identity to the
 CC above sequences. The nucleic acid is useful for preparing a composition
 CC for diagnosing or treating psoriasis in a mammal. This sequence
 CC corresponds to one of the polypeptides of the invention.
 XX
 SQ Sequence 2000 AA;
 XX

Query Match 17.1%; Score 7; DB 8; Length 2000;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 APERKER 34
 Db 1333 APERKER 1339

RESULT 9
 AAM39043
 ID AAM39043 standard; protein; 4618 AA.
 XX
 AC AAM39043;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 2188.
 XX
 KW Human; nootropic; immunosuppressant; cyrostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US034263.
 XX
 PR 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-00488725.
 PR 25-APR-2000; 2000US-00552317.
 PR 20-JUN-2000; 2000US-00598042.
 PR 19-JUL-2000; 2000US-00620312.

PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
 PI Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR N-PSDB; AAI58199.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.
 XX
 PS Example 4, SEQ ID NO 2188; 10078pp, English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and the
 CC encoded polypeptides (AAM38642-AA42213) with nootropic.
 CC immunosuppressant and cyrostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification
 XX
 SQ Sequence 4618 AA;
 XX

Query Match 17.1%; Score 7; DB 4; Length 4618;
 Best Local Similarity 100.0%; Pred. No. 4.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 APERKER 34
 Db 1544 APERKER 1550

RESULT 10
 ADL46346
 ID ADL46346 standard; peptide; 21 AA.
 XX
 AC ADL46346;
 XX
 DT 20-MAY-2004 (first entry)
 XX
 DE D-alanine:D-alanine adding enzyme epitope #1.
 XX
 KW epitope; antibacterial;
 KW UDP-N-acetylglucosamine 1-carboxyvinyl transferase-1;
 KW CMP:UMP-3-deoxy-D-manno-octulosonate transferase;
 KW UDP-N-acetylmuramylalanyl-D-glutamate-2-6-diaminopimelate ligase;
 KW D-alanine-D-alanine adding enzyme; D-alanine-D-alanine ligase;
 KW UDP-N-acetylpyruvoylglucosamine reductase;
 KW UDP-N-acetylglucosamine pyrophosphorylase;
 KW UDP-N-acetylmuramylalanine-D-glutamate ligase;
 KW DP-N-acetylmuramate:alanine ligase; separate semialdehyde dehydrogenase;
 KW UDP-N-acetylmuramylalanyl-D-glutamate; X-ray diffraction analysis;
 KW enzyme.
 XX
 OS Pseudomonas aeruginosa.
 XX
 PN WO2003087353-A2.
 XX
 PD 23-OCT-2003.

XX 08-APR-2003; 2003WO-CA000481.
PF
XX 08-APR-2002; 2002US-0370899P.
PR 08-APR-2002; 2002US-0370915P.
PR 09-APR-2002; 2002US-0371107P.
PR 09-APR-2002; 2002US-0371185P.
PR 31-MAY-2002; 2002US-0385426P.
PR 06-JUN-2002; 2002US-0386283P.
PR 01-AUG-2002; 2002US-0400348P.
PR 06-NOV-2002; 2002US-0424395P.
PR 08-NOV-2002; 2002US-0425200P.
PR 24-DEC-2002; 2002US-0436345P.
PR 24-DEC-2002; 2002US-0436349P.
PR 26-DEC-2002; 2002US-0436568P.
PR 27-DEC-2002; 2002US-0436734P.
PR 27-DEC-2002; 2002US-0436885P.
PR 27-DEC-2002; 2002US-0436893P.
PR 27-DEC-2002; 2002US-0436900P.
PR 30-DEC-2002; 2002US-0437013P.
XX
PA (AEFI-) AFFINIDM PHARM INC.
PI Edwards A, Dharamsi A, Vedadi M, Domagala M, Houston S, Awrey D;
PI Beattie B, Mansoury K, Ouyang H, Vallee F, Richards D, Nehery K;
PI Virag C, Buzaditja K, Pinder B, Alam MZ, Tai M, Canadian V;
PI Kanagarajah D, Thalakada R;
XX
DR WPI; 2003-865361/80.
XX
PT New recombinant bacterial enzymes involved in cell membrane biogenesis,
PT useful for designing potential antibacterial agents.
XX
PS Disclosure; SEQ ID NO 64; 407pp; English.
XX
XX The invention relates to isolated, recombinant polypeptides (I) that have
CC at least one activity of specified bacterial enzymes involved in cell
CC membrane biogenesis. (I) are: UDP-N-acetylglucosamine 1-carboxyvinyl
CC transferase-1 of *Streptococcus pneumoniae* (S.p.), *Pseudomonas aeruginosa*
CC (P.a.) or *Staphylococcus aureus* (S.a.); CTP:CM-3-deoxy-D-manno-
CC octulosonate transferase of *Escherichia coli* (E.c.) or *Haemophilus*
CC influenzae (H.i.); UDP-N-acetylmuramylalanyl-D-glutamate-2,6-
CC diaminopimelate ligase of P.a.; D-alanine-D-alanine adding enzyme of S.a.
CC or P.a.; D-alanine-D-alanine ligase of *Enterococcus faecalis* (E.f.); UDP-N-
CC -acetylpyruvoylglucosamine reductase of P.a. or H.i.; UDP-N-
CC acetylglucosamine pyrophosphorylase of E.f., H.i. or S.a.; DP-N-
CC acetylmuramoylalanine-D-glutamate ligase of E.f. or H.i.; DP-N-
CC acetylmuramate:alanine ligase of E.c.; and aspartate semialdehyde
CC dehydrogenase of H.i. and UDP-N-acetylmuramoylalanyl-D-glutamate (sic) of
CC H.i. Crystalline (I) are used to determine (by X-ray diffraction
CC analysis) the structural coordinates of (I), and these then used to
CC design modulators of (I), potential therapeutic agents for treating
CC diseases caused by the specified bacteria. This sequence represents an
CC epitope from one of the proteins of the invention.
XX
SQ Sequence 21 AA;
Query Match 14.6%; Score 6; DB 7; Length 21;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 15 PFVPL 20
Db 14 PFVPL 19

XX 07-NOV-2001 (first entry)
DT
XX Human immune/haematopoietic antigen SEQ ID NO:16666.
DB Human immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
XX 09-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US001354.
PF
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227709P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229350P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0232081P.
PR 14-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232388P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.

```

PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 27-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249287P.
PR 17-NOV-2000; 2000US-0249289P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.

```

```

PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX N-PSDB; AAK62054.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides;
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Claim 11; SEQ ID NO 1686; 3071bp + Sequence listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting the
XX nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially,
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX represent sequences used in the exemplification of the present invention
XX
XX Sequence 32 AA:
XX
XX Query Match 14.6%; Score 6; DB 4; Length 32;
XX Best Local Similarity 100.0%; Pred. No. 97;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 35 KNTVIGS 40
XX Db 12 KNTVIGS 17
XX
XX RESULT 12
XX AAM20415
XX ID AAM20415 standard; protein; 47 AA.
XX AC AAM20415;
XX XX
XX DT 12-OCT-2001 (first entry)
XX DE Peptide #6849 encoded by probe for measuring cervical gene expression.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer.
XX OS Homo sapiens.
XX XX
XX PN WO200157278-A2.
XX XX
XX PD 09-AUG-2001.
XX XX
XX PF 30-JAN-2001; 2001WO-US000670.
XX XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.

```

```
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human cervical epithelial cells.
XX
XX Claim 27; SEQ ID NO 25241; 487bp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENPs: see A1110068-A1128459). The present sequence is a peptide encoded
XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 47 AA;
SQ
Query Match 14.6%; Score 6; DB 4; Length 47;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 30 ERKEEK 35
Db 33 ERKEEK 38
RESULT 13
ABB41137
ID ABB41137 standard; peptide; 47 AA.
XX
XX ABB41137;
XX
XX 04-FEB-2002 (first entry)
XX
XX Peptide #8643 encoded by human foetal liver single exon probe.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
XX Homo sapiens.
XX
XX WO200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000669.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human fetal liver.
```

```
XX
XX Claim 27; SEQ ID NO 33772; 639bp + Sequence Listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for measuring
XX human gene expression in a sample derived from human foetal liver. The
XX single exon nucleic acid probes may be used for predicting, measuring and
XX displaying gene expression in samples derived from human fetal liver. The
XX present sequence is a peptide encoded by a single exon nucleic acid probe
XX of the invention. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 47 AA;
SQ
Query Match 14.6%; Score 6; DB 4; Length 47;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 30 ERKEEK 35
Db 33 ERKEEK 38
RESULT 14
AAM34913
ID AAM34913 standard; protein; 47 AA.
XX
XX AAM34913;
XX
XX 17-OCT-2001 (first entry)
XX
XX Peptide #8950 encoded by probe for measuring placental gene expression.
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000663.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX
XX Claim 27; SEQ ID NO 35182; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs:
XX see A113115-A1157546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders
XX
XX Sequence 47 AA;
SQ
Query Match 14.6%; Score 6; DB 4; Length 47;
```

Best Local Similarity 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

Search completed: February 1, 2005, 15:20:10
Job time : 71.7778 secs

QY 30 ERKEEK 35
33 ERKEEK 38

RESULT 15

ABB25176
ID ABB25176 standard; protein; 47 AA.

XX ABB25176;

DT 23-JAN-2002 (first entry)

DE Protein #7175 encoded by probe for measuring heart cell gene expression.

XX Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;

KW congenital heart disease.

XX OS Homo sapiens.

XX WO200157274-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000666.

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207455P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX MPI; 2001-488899/53.

PT Single exon nucleic acid probes for analyzing gene expression in human

PT hearts.

PS Claim 15; SEQ ID NO 26946; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for

XX measuring human gene expression in a sample derived from human heart (see

XX ABA21535-ABA41305). The present sequence is a protein encoded by one such

XX probe. The probes may be used for predicting, measuring and displaying

XX gene expression in samples derived from the human heart via microarrays.

XX By measuring gene expression, the probes are useful for predicting,

XX diagnosing, grading, staging, monitoring and prognosing diseases of the

XX human heart and vascular system e.g. cardiovascular disease,

XX hypertension, cardiac arrhythmias and congenital heart disease. Note: The

XX sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 47 AA;

QY 30 ERKEEK 35

DB 33 ERKEEK 38

Query Match 14.6%; Score 6; DB 4; Length 47;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 ERKEEK 35
DB 33 ERKEEK 38

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 15:09:39 ; Search time 59.7778 Seconds
(without alignments)
247.799 Million cell updates/sec

Title: SEQ5ASN1694

Perfect score: 41
Sequence: 1 LKLCNKGAFNPDPPVPL.....TAVXLIAPERKEKNVLGSA 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1608061 seqs, 361289386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1311380

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	1149	17 US-10-128-558-167	Sequence 167, App
2	7	17.1	61	16 US-10-767-701-52614	Sequence 52614, A
3	7	17.1	75	15 US-10-424-599-224449	Sequence 224449, A
4	7	17.1	79	15 US-10-424-599-157348	Sequence 157348, A
5	7	17.1	93	15 US-10-424-599-198515	Sequence 198515, A
6	7	17.1	98	15 US-10-424-599-271135	Sequence 271135, A
7	7	17.1	119	15 US-10-424-599-283405	Sequence 283405, A
8	7	17.1	169	16 US-10-767-701-53909	Sequence 53909, A
9	7	17.1	186	15 US-10-424-599-279781	Sequence 279781, A
10	7	17.1	232	15 US-10-424-599-157336	Sequence 157336, A
11	7	17.1	265	15 US-10-424-599-157334	Sequence 157334, A
12	7	17.1	308	15 US-10-282-122A-52352	Sequence 52352, A
13	7	17.1	380	14 US-10-369-493-924	Sequence 924, App

14	7	17.1	416	16 US-10-437-963-109626	Sequence 109626, A
15	7	17.1	574	16 US-10-437-963-109624	Sequence 109624, A
16	6	14.6	51	16 US-10-437-963-152060	Sequence 152060, A
17	6	14.6	47	9 US-09-864-761-40474	Sequence 40474, A
18	6	14.6	50	9 US-09-764-877-1063	Sequence 1063, Ap
19	6	14.6	50	15 US-10-242-515-1063	Sequence 1063, Ap
20	6	14.6	50	17 US-10-425-115-253833	Sequence 253833, A
21	6	14.6	53	17 US-10-425-115-285594	Sequence 285594, A
22	6	14.6	55	15 US-10-424-599-254164	Sequence 254164, A
23	6	14.6	61	13 US-10-002-344A-231	Sequence 231, App
24	6	14.6	62	15 US-10-282-122A-48051	Sequence 48051, A
25	6	14.6	64	17 US-10-425-115-335280	Sequence 335280, A
26	6	14.6	67	14 US-10-156-761-8250	Sequence 8250, Ap
27	6	14.6	67	15 US-10-424-599-218416	Sequence 218416, A
28	6	14.6	68	16 US-10-767-701-61219	Sequence 61219, A
29	6	14.6	69	17 US-10-425-115-264251	Sequence 264251, A
30	6	14.6	70	9 US-09-860-192-49	Sequence 49, Appl
31	6	14.6	70	14 US-10-427-348-49	Sequence 49, Appl
32	6	14.6	71	15 US-10-424-599-238496	Sequence 238496, A
33	6	14.6	72	17 US-10-425-115-264469	Sequence 264469, A
34	6	14.6	74	15 US-10-424-599-187518	Sequence 187518, A
35	6	14.6	79	17 US-10-425-115-293677	Sequence 293677, A
36	6	14.6	79	17 US-10-425-115-308354	Sequence 308354, A
37	6	14.6	85	17 US-10-425-115-269295	Sequence 269295, A
38	6	14.6	90	16 US-10-437-963-146939	Sequence 146939, A
39	6	14.6	92	15 US-10-424-599-171593	Sequence 171593, A
40	6	14.6	92	17 US-10-425-115-369145	Sequence 369145, A
41	6	14.6	95	15 US-10-424-599-193871	Sequence 193871, A
42	6	14.6	95	15 US-10-296-115-1451	Sequence 1451, Ap
43	6	14.6	98	17 US-10-425-115-249042	Sequence 249042, A
44	6	14.6	98	17 US-10-425-115-319185	Sequence 319185, A
45	6	14.6	101	16 US-10-437-963-170845	Sequence 170845, A

ALIGNMENTS

RESULT 1
US-10-128-558-167
Sequence 167, Application US/10128558
Publication No. US20040219521A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Wang, Zhiwei
APPLICANT: Wong, Gezhi
APPLICANT: Boyle, Bryan J
APPLICANT: Drmanac, Radco J
TITLE OF INVENTION: Novel Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 812A
CURRENT APPLICATION NUMBER: US/10/128,558
CURRENT FILING DATE: 2002-04-22
PRIOR APPLICATION NUMBER: US 60/339,453
PRIOR FILING DATE: 2001-12-11
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02823
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: PCT/US01/03800
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 09/515,126
PRIOR FILING DATE: 2000-02-28
Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 412
 SOFTWARE: PL_genes Version 6.0
 SEQ ID NO 167
 LENGTH: 1149
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-128-558-167

Query Match 100.0%; Score 41; DB 17; Length 1149;
 Best Local Similarity 100.0%; Pred. No. 4,3e-33;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LKLCNKGANPPFVPLNTAVKLTPERKEKNVLSA 41
 DB 732 LKLCNKGANPPFVPLNTAVKLTPERKEKNVLSA 772

RESULT 2

US-10-767-701-52614
 Sequence 52614, Application US/10767701
 Publication No. US20040172684A1
 GENERAL INFORMATION:
 APPLICANT: Kovalic, David K.
 APPLICANT: Zhou, Yihua
 APPLICANT: Cao, Yongwei
 TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
 FILE REFERENCE: 38-21(53223)B
 CURRENT APPLICATION NUMBER: US/10/767,701
 CURRENT FILING DATE: 2004-01-29
 NUMBER OF SEQ ID NOS: 63128
 SEQ ID NO 52614
 LENGTH: 61
 TYPE: PRT
 ORGANISM: Sorghum bicolor
 FEATURE:
 OTHER INFORMATION: Clone ID: 12691726.pep
 US-10-767-701-52614

Query Match 17.1%; Score 7; DB 16; Length 61;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 APERKEE 34
 DB 19 APERKEE 25

RESULT 3

US-10-424-599-224449
 Sequence 224449, Application US/10424599
 Publication No. US20040031072A1
 GENERAL INFORMATION:
 APPLICANT: La Rosa, Thomas J
 APPLICANT: Kovalic, David K
 APPLICANT: Zhou, Yihua
 APPLICANT: Cao, Yongwei
 TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
 FILE REFERENCE: 38-21(53223)B
 CURRENT APPLICATION NUMBER: US/10/424,599
 CURRENT FILING DATE: 2003-04-28
 NUMBER OF SEQ ID NOS: 285684
 SEQ ID NO 224449
 LENGTH: 75
 TYPE: PRT
 ORGANISM: Glycine max
 FEATURE:
 OTHER INFORMATION: Clone ID: PAT_MRT3847_44706C.1.pep
 US-10-424-599-224449

Query Match 17.1%; Score 7; DB 15; Length 75;
 Best Local Similarity 100.0%; Pred. No. 21;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 PERKEEK 35
 DB 4 PERKEEK 10

RESULT 4

US-10-424-599-157348
 Sequence 157348, Application US/10424599
 Publication No. US20040031072A1
 GENERAL INFORMATION:
 APPLICANT: La Rosa, Thomas J
 APPLICANT: Kovalic, David K
 APPLICANT: Zhou, Yihua
 APPLICANT: Cao, Yongwei
 TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
 FILE REFERENCE: 38-21(53223)B
 CURRENT APPLICATION NUMBER: US/10/424,599
 CURRENT FILING DATE: 2003-04-28
 NUMBER OF SEQ ID NOS: 285684
 SEQ ID NO 157348
 LENGTH: 79
 TYPE: PRT
 ORGANISM: Glycine max
 FEATURE:
 NAME/KEY: unsure
 LOCATION: (1)-(79)
 OTHER INFORMATION: unsure at all Xaa locations
 FEATURE:
 OTHER INFORMATION: Clone ID: PAT_MRT3847_113105C.1.pep
 US-10-424-599-157348

Query Match 17.1%; Score 7; DB 15; Length 79;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 PERKEEK 35
 DB 25 PERKEEK 31

RESULT 5

US-10-424-599-198515
 Sequence 198515, Application US/10424599
 Publication No. US20040031072A1
 GENERAL INFORMATION:
 APPLICANT: La Rosa, Thomas J
 APPLICANT: Kovalic, David K
 APPLICANT: Zhou, Yihua
 APPLICANT: Cao, Yongwei
 TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
 FILE REFERENCE: 38-21(53223)B
 CURRENT APPLICATION NUMBER: US/10/424,599
 CURRENT FILING DATE: 2003-04-28
 NUMBER OF SEQ ID NOS: 285684
 SEQ ID NO 198515
 LENGTH: 93
 TYPE: PRT
 ORGANISM: Glycine max
 FEATURE:
 OTHER INFORMATION: Clone ID: PAT_MRT3847_21283C.1.pep
 US-10-424-599-198515

Query Match 17.1%; Score 7; DB 15; Length 93;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 PERKEEK 35
 DB 64 PERKEEK 70

```
RESULT 6
US-10-424-599-271135
; Sequence 271135, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 271135
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_86854C.1.pep
US-10-424-599-271135

Query Match      17.1%; Score 7; DB 15; Length 98;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 LKLLCKN 7
        |||||
Db      70 LKLLCKN 76

RESULT 7
US-10-424-599-283405
; Sequence 283405, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 283405
; LENGTH: 119
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(119)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_97938C.1.pep
US-10-424-599-283405

Query Match      17.1%; Score 7; DB 15; Length 119;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      32 KEKKNVL 38
        |||||
Db      67 KEKKNVL 73

RESULT 8
US-10-767-701-53909
; Sequence 53909, Application US/10767701
```

```
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 53909
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 13784602.pep
US-10-767-701-53909

Query Match      17.1%; Score 7; DB 16; Length 169;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      28 APERKEE 34
        |||||
Db      29 APERKEE 35

RESULT 9
US-10-424-599-279781
; Sequence 279781, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 279781
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(186)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_94664C.1.pep
US-10-424-599-279781

Query Match      17.1%; Score 7; DB 15; Length 186;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      30 ERKEEKN 36
        |||||
Db      73 ERKEEKN 79

RESULT 10
US-10-424-599-157336
; Sequence 157336, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
```

```

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 157336
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_113095C.1.pep
US-10-424-599-157336

Query Match          17.1%; Score 7; DB 15; Length 232;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      32 KEEKNVL 38
Db      180 KEEKNVL 186

RESULT 11
US-10-424-599-157334
; Sequence 157334, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 157334
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_113093C.1.pep
US-10-424-599-157334

Query Match          17.1%; Score 7; DB 15; Length 265;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      32 KEEKNVL 38
Db      213 KEEKNVL 219

RESULT 12
US-10-282-122A-52352
; Sequence 52352, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forayth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

```

```

; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52352
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-282-122A-52352

Query Match          17.1%; Score 7; DB 15; Length 308;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      35 KNVLGSA 41
Db      298 KNVLGSA 304

RESULT 13
US-10-369-493-924
; Sequence 924, Application US/10369493
; Publication No. US20030231675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 924
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Archaeoglobus fulgidus
US-10-369-493-924

Query Match          17.1%; Score 7; DB 14; Length 380;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      18 PVLNTAV 24
Db      120 PVLNTAV 126

```

RESULT 14

US-10-437-963-109626
; Sequence 109626, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 109626
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(416)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_13766C.1.pep
US-10-437-963-109626

Query March

Best Local Similarity 17.1%; Score 7; DB 16; Length 416;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 ERKEKN 36
|||

Db 296 ERKEKN 302

RESULT 15

US-10-437-963-109624
; Sequence 109624, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 109624
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_13764C.1.pep
US-10-437-963-109624

Query March

Best Local Similarity 17.1%; Score 7; DB 16; Length 574;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 ERKEKN 36

Db 184 ERKEKN 190
|||||

Search completed: February 1, 2005, 15:44:47
Job time : 60.7778 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:55:01 / Search time 13.5556 Seconds
(without alignments)
291.016 Million cell updates/sec

Title: SE05ASN1694

Perfect score: 41
Sequence: 1 LKLLCKNFGAENPDPPFVVL.....TAVKLIAPERKEKNVLGSA 41Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size: 0

Total number of hits satisfying chosen parameters: 279530

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: PIR 79:*

1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	17.1	330	2 T35360	probable membrane
2	7	17.1	380	2 B69275	3-ketoadyl-CoA thi
3	7	17.1	435	2 B72418	conserved hypotnet
4	7	17.1	662	2 D54078	methyl-accepting c
5	7	17.1	1756	1 BVECAI	DNA helicase I (EC
6	7	17.1	3566	1 A40701	tenascin-X precurs
7	6	14.6	45	2 H61526	hypothetical prote
8	6	14.6	77	2 CT0306	conserved hypotnet
9	6	14.6	105	2 B70181	conserved hypotnet
10	6	14.6	109	2 B61180	protein-cysteine-p
11	6	14.6	114	2 T34582	hypothetical prote
12	6	14.6	125	2 S04503	pancreatic ribonuc
13	6	14.6	139	2 A29297	hypothetical prote
14	6	14.6	141	2 B38286	hypothetical prote
15	6	14.6	142	2 T51479	hypothetical prote
16	6	14.6	144	2 A70411	small heat shock p
17	6	14.6	148	2 T49397	hypothetical prote
18	6	14.6	154	2 S03171	interleukin-7 prec
19	6	14.6	157	2 G72864	Acortf-118 protein
20	6	14.6	187	2 S26139	signaling protein
21	6	14.6	187	2 A48901	signaling protein
22	6	14.6	188	2 B83080	beta-lactamase exp
23	6	14.6	192	2 T40605	hypothetical prote
24	6	14.6	197	2 S78187	ymf39 protein - Re
25	6	14.6	199	2 S75344	hypothetical prote
26	6	14.6	215	2 T23482	hypothetical prote
27	6	14.6	238	2 A31417	prolactin-related
28	6	14.6	249	2 H64368	cobalamin biosynth
29	6	14.6	266	2 T19236	hypothetical prote

30	6	14.6	277	2 T40033	probable microchond
31	6	14.6	279	2 T48013	hypothetical prote
32	6	14.6	289	2 AD1086	ATP synthase gamma
33	6	14.6	295	2 S76136	hypothetical prote
34	6	14.6	298	2 T31518	hypothetical prote
35	6	14.6	302	2 A84331	hypothetical prote
36	6	14.6	305	2 A69295	UDP-glucose 4-epim
37	6	14.6	315	2 T26422	hypothetical prote
38	6	14.6	316	2 AF3138	conserved hypotnet
39	6	14.6	316	2 E98149	hypothetical prote
40	6	14.6	336	2 T34452	hypothetical prote
41	6	14.6	340	2 T30121	hypothetical prote
42	6	14.6	341	2 G02516	aripatin 2 - human
43	6	14.6	343	2 H84607	hypothetical prote
44	6	14.6	348	2 S44628	f22b7.1 protein -
45	6	14.6	359	2 T18667	hypothetical prote

ALIGNMENTS

RESULT 1

T35360 Probable membrane protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004

C:Accession: T35360

R:Murphy, L.; Harris, D.; James, K.D.; Parthill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, June 1999

A:Reference number: 221576

A:Accession: T35360

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-330 <MIR>

A:Cross-references: UNIPROT:Q9YAL0; EMBL:AL079348; PIDD:CAM45459.1; GSPDB:GN00070; SCOPED

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOPED:SC6673.03

Query Match

Best Local Similarity 17.1%; Score 7; DB 2; Length 330;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 9 GAENPD 15

Db 109 GAENPD 115

RESULT 2

B69275 3-ketoadyl-CoA thiolase (acsB-6) homolog - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004

C:Accession: B69275

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson

; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirsness, E.F.

; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.I.

Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo

A:Reference number: A69250; MIMD:98049343; PMID:9389475

A:Accession: B69275

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-380 <KLE>

A:Cross-references: UNIPROT:Q30037; GB:AE001092; GB:AE000782; NID:g2689415; PIDD:AAB9102;

Query Match

Best Local Similarity 17.1%; Score 7; DB 2; Length 380;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 18 PVLNTAV 24


```

A/Molecule type: DNA
A/Residues: 'M',1326-1332,'P',1334-1335,'T',1337-1352,'G',1354-1375,'G',1377-1385,'A',13
1717-1719,'R',1721-1726,'V',1728-1738,'W',1740-1756<EURL>
A/Cross-references: EMBL:AF074613; PDB:1AC70166.1
A/Experimental source: strain EOL933; serotype O157:H7
C/Comment: This is one of the proteins that control the transfer of F plasmid.
C/Comment: There seems to be some variation or disagreement regarding the translation s
C/Genetics:
A/Genome: tral
A/Genome: plasmid
C/Superfamily: helicase I
C/Keywords: ATP; DNA binding; DNA repair; DNA replication; F pilin formation; hydrolase;
F/992-999/Region: nucleotide-binding motif A (P-loop)

Query Match      17.1%; Score 7; DB 1; Length 1756;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      17 VPVINTA 23
Db      1261 VPVINTA 1267

RESULT 6
A40701
C:Species: Homo sapiens (man)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C/Accession: A40701; A33725; C42175
R/Bristow, J.; Tee, M.K.; Gitelman, S.E.; Mellon, S.H.; Miller, W.L.
J. Cell Biol. 122, 265-278, 1993
A/Title: Tenascin-X: a novel extracellular matrix protein encoded by the human XB gene c
A/Reference number: A40701; MUID:93300909; PMID:7686164
A/Accession: A40701
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-3566<BR>
A/Cross-references: UNIPROT:P2105; UNIPROT:Q9UC11; EMBL:X71937
R/McCrel, Y.; Bristow, J.; Gitelman, S.E.; Miller, W.L.
Proc. Natl. Acad. Sci. U.S.A. 86, 6582-6586, 1989
A/Title: Transcript encoded on the opposite strand of the human steroid 21-hydroxylase/c
A/Reference number: A33725; MUID:89367293; PMID:2475872
A/Accession: A33725
A/Molecule type: mRNA
A/Residues: 2748-3199,'V',3201-3298,'E',3299-3314,'G',3316-3566<MOR>
A/Cross-references: GEM:W5813; NID:G183069; PDB:1AA35884.1; PDB:G183070
R/Matsuoto, K.; Arai, M.; Ishihara, N.; Ando, A.; Inoko, H.; Ikemura, T.
Genomics 12, 485-491, 1992
A/Title: Cluster of fibronectin type III repeats found in the human major histocompatib
enacin.
A/Reference number: A42175; MUID:92217969; PMID:1373119
A/Accession: C42175
A/Molecule type: DNA
A/Residues: 1849-1936<MAT>
A/Experimental source: clone 3.9Kf3-1
A/Note: sequence extracted from NCBI backbone (NCBI:95594)
C/Genetics:
A/Genes: GDB:TNXA; D6S103E; TNX; XA; XB
A/Cross-references: GDB:568487; OMIM:600261
A/Map position: 6p21.3-6p21.3
C/Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin ty
C/Keywords: extracellular matrix; glycoprotein
F/435-461/Domain: EGF homology <EGF>
F/748-828/Domain: fibronectin type III repeat homology <3F1>
F/828-856/Domain: fibronectin type III repeat homology #status atypical <3F2>
F/873-953/Domain: fibronectin type III repeat homology <3F3>
F/975-1055/Domain: fibronectin type III repeat homology <3F4>
F/1078-1158/Domain: fibronectin type III repeat homology <3F5>
F/1167-1247/Domain: fibronectin type III repeat homology <3F6>
F/1248-1317/Domain: fibronectin type III repeat homology #status atypical <3F7>
F/1333-1403/Domain: fibronectin type III repeat homology <3F8>
F/1412-1492/Domain: fibronectin type III repeat homology <3F9>
F/1510-1590/Domain: fibronectin type III repeat homology <3F10>

```

```

F/1618-1676/Domain: fibronectin type III repeat homology #status atypical <3F11>
F/1678-1749/Domain: fibronectin type III repeat homology #status atypical <3F12>
F/1761-1831/Domain: fibronectin type III repeat homology <3F13>
F/1849-1929/Domain: fibronectin type III repeat homology <3F14>
F/1955-2035/Domain: fibronectin type III repeat homology <3F15>
F/2061-2141/Domain: fibronectin type III repeat homology <3F16>
F/2167-2246/Domain: fibronectin type III repeat homology <3F17>
F/2274-2354/Domain: fibronectin type III repeat homology <3F18>
F/2382-2462/Domain: fibronectin type III repeat homology <3F19>
F/2488-2568/Domain: fibronectin type III repeat homology <3F20>
F/2584-2664/Domain: fibronectin type III repeat homology <3F21>
F/2677-2757/Domain: fibronectin type III repeat homology <3F22>
F/2771-2851/Domain: fibronectin type III repeat homology <3F23>
F/2878-2958/Domain: fibronectin type III repeat homology <3F24>
F/2977-3067/Domain: fibronectin type III repeat homology #status atypical <3F25>
F/3078-3159/Domain: fibronectin type III repeat homology <3F26>
F/3167-3247/Domain: fibronectin type III repeat homology <3F27>
F/3255-3334/Domain: fibronectin type III repeat homology <3F28>
F/3349-3557/Domain: fibrinogen beta/gamma homology <F86>

Query Match      17.1%; Score 7; DB 1; Length 3566;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      28 APERKEE 34
Db      961 APERKEE 967

RESULT 7
H81526
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C/Accession: H81526
R/Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A/Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
A/Reference number: H81500; MUID:20150255; PMID:10684935
A/Accession: H81526
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-45<RBA>
A/Cross-references: UNIPROT:Q9Y1M1; GB:AF002247; GB:AB002161; NID:G7189792; PDB:1AF38665
A/Experimental source: strain AR39, HL cells
C/Genetics:
A/Genes: CP0881

Query Match      14.6%; Score 6; DB 2; Length 45;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      31 RKEEKN 36
Db      25 RKEEKN 30

RESULT 8
C70306
C:Species: Aquifex aeolicus
C/Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C/Accession: C70306
R/Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A/Reference number: A70300; MUID:98196666; PMID:9537320
A/Accession: C70306
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-77<AOF>

```

A:Cross-references: UNIPROT:O66478; GB:AE000672; NID:g2982810; PIDN:AA06451.1; PID:g298
 A:Experimental source: strain VFS
 C:Genetics:
 A:Gene: aq_064c
 C:Superfamily: conserved hypothetical secreted protein HP0320

Query Match 14.6%; Score 6; DB 2; Length 77;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 ERKEK 35
 |||||
 DB 63 ERKEK 68

RESULT 9

B70181
 conserved hypothetical protein BB0651 - Lyme disease spirochete

C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004

C:Accession: B70181

R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vgnt,
 ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997
 A:Author: Smith, H.O.; Venter, J.C.

A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943; PMID:9403685

A:Accession: B70181

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-105 <KLB>

A:Cross-references: UNIPROT:O51595; GB:AE001166; GB:AE000783; NID:g2688571; PIDN:AA06695
 A:Experimental source: strain B31

C:Superfamily: yajc protein

Query Match 14.6%; Score 6; DB 2; Length 105;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 RKEKN 36
 |||||
 DB 43 RKEKN 48

RESULT 10

B61180

protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type epsilon - mouse (fragment)
 N:Alternate names: protein-tyrosine-phosphatase PTPY8

C:Species: Mus musculus (house mouse)

C:Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 09-Jul-2004

C:Accession: B61180; S40283

R:Yi, T.; Cleveland, J.L.; Ihle, J.N.

Blood 78, 2222-2228, 1991

A:Title: Identification of novel protein tyrosine phosphatases of hematopoietic cells by
 A:Reference number: A61180; MUID:92032882; PMID:1932742

A:Accession: B61180

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-109 <YTA>

A:Cross-references: UNIPROT:P49446

R:Hendriks, W.; Brugman, C.; Zeuwen, P.; Schepens, J.; Wieringa, B.
 submitted to the EMBL Data Library, June 1993

A:Description: Assessment of the expression levels of murine protein-tyrosine phosphatase

A:Reference number: S40280

A:Accession: S40283

A:Molecule type: mRNA

A:Residues: 1-30, T, 32-109 <HEN>

A:Cross-references: EMBL:Z23052; NID:9438141; PIDN:CA060587.1; PID:9438142
 C:Superfamily: protein-tyrosine-phosphatase, receptor type alpha; leukocyte common antige
 n; Keywords: phosphoric monoester hydrolase; receptor; transmembrane protein; tyrosine-ph
 C:Keywords: phosphoric monoester hydrolase; receptor; transmembrane protein; tyrosine-ph
 F:1-109/Domain: leukocyte common antigen cytosolic domain homology (fragment) <LAC>

Query Match 14.6%; Score 6; DB 2; Length 109;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 ERKEK 35
 |||||
 DB 14 ERKEK 19

RESULT 11

T34582

hypothetical protein SC10A5.15 SC10A5.15 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004

C:Accession: T34582

R:Murphy, U.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, January 1998

A:Reference number: Z21548

A:Accession: T34582

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-114 <MCR>

A:Cross-references: UNIPROT:O54106; EMBL:AL021529; PIDN:CA016447.1; GSPDB:GND00070; SCODE
 A:Experimental source: strain A3(2)

C:Genetics:
 A:Gene: SCODEB:SC10A5.15

Query Match 14.6%; Score 6; DB 2; Length 114;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 AVKLIA 28
 |||||
 DB 107 AVKLIA 112

RESULT 12

S04503

pancreatic ribonuclease (EC 3.1.27.5) - Ehrenberg's mole-rat

C:Species: Spalax leucodon ehrenbergi (Ehrenberg's mole-rat)

C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 09-Jul-2004

C:Accession: S04503; A32505

R:Schueller, C.; Neuteboom, B.; Wuebbels, G.H.; Beintema, J.J.; Nevo, E.
 Biol. Chem. Hoppe-Seyler 370, 583-589, 1989

A:Title: The amino-acid sequence of pancreatic ribonuclease from the mole rat Spalax ehr

A:Reference number: S04503; MUID:89374807; PMID:2673297

A:Accession: S04503

A:Molecule type: protein

A:Residues: 1-125 <SCH>

A:Cross-references: UNIPROT:P16414

A>Note: the source is designated as Spalax ehrenbergi

C:Superfamily: pancreatic ribonuclease

C:Keywords: glycoprotein; hydrolase

F:1,42,120/Active site: His, Lys, His #status predicted

F:27-85,41-96,59-111,66-73/Disulfide bonds: #status predicted

Query Match 14.6%; Score 6; DB 2; Length 125;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 DPFVPV 19
 |||||
 DB 114 DPFVPV 119

RESULT 13

AD2997

hypothetical protein Atus582 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004

C:Accession: AD2997

R:Wood, D.W.; Seubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan

; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 eter, E.W.
 A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A/Reference number: AB2577; MUID:21608550; PMID:11743193
 A/Accession: AD2997
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-139 <KIR>
 A/Cross-references: UNIPROT:Q8U9Z3; GB:AE008689; PIDN:AAL44394.1; PID:q17741992; GSPDB:C
 A/Experimental source: strain C58 (Dupont)
 C/Genetics:
 A/Gene: Atus582
 A/Map position: linear chromosome

Query Match 14.6%; Score 6; DB 2; Length 139;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 15 PFVPVL 20
 Db 96 PFVPVL 101

RESULT 14
 E98286
 hypothetical protein AGR_L_2491 [imported] - Agrobacterium tumefaciens (strain C58, Cere
 C/Species: Agrobacterium tumefaciens
 C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
 C/Accession: E98286
 R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tun
 A/Reference number: A97359; MUID:21608551; PMID:11743194
 A/Accession: E98286
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-141 <KIR>
 A/Cross-references: UNIPROT:Q8U9Z3; GB:AE007870; PIDN:AAK89815.1; PID:q15159745; GSPDB:C
 C/Genetics:
 A/Gene: AGR_L_2491
 A/Map position: linear chromosome

Query Match 14.6%; Score 6; DB 2; Length 141;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 PFVPVL 20
 Db 98 PFVPVL 103

RESULT 15
 T51479
 hypothetical protein T21H19_10 - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 09-Jul-2004
 C/Accession: T51479
 R/Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew
 submitted to the Protein Sequence Database, August 2000
 A/Reference number: 225394
 A/Accession: T51479
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-142 <SAT>
 A/Cross-references: UNIPROT:Q9ULF19; EMBL:AL391148
 A/Experimental source: cultivar Columbia; BAC clone T21H19
 C/Genetics:
 A/Map position: 5
 A/Introns: 22/3; 76/3; 111/2
 A/Note: T21H19_10

Query Match 14.6%; Score 6; DB 2; Length 142;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 FVPVLN 21
 Db 100 FVPVLN 105

Search completed: February 1, 2005, 15:33:00
 Job time : 13.5556 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:54:06 ; Search time 70.5556 Seconds

(without alignments)
334.352 Million cell updates/sec

Title: SEQ5ASN1694

Perfect score: 41

Sequence: 1 LKLLCKNFGAENPDFFVPL.....TAVGLIAPERKEKNVLGSA 41

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1825181 seqs, 575374646 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1810864

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Uniprot_02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	41	100.0	2144 1 BP28_HUMAN	Q9H583 homo sapien
2	20	48.8	897 2 Q8N7L7	Q8N7L7 homo sapien
3	20	48.8	958 1 BP28_MACPA	Q9Gm4 macaca fasc
4	20	48.8	1106 2 Q6P157	Q6P157 homo sapien
5	20	48.8	1106 2 AAH65205	AAH65205 homo sap
6	8	19.5	181 2 Q87ZNS	Q87ZNS pseudomona
7	8	19.5	585 2 Q8BH7	Q8BH7 mus musculi
8	8	19.5	920 2 Q8CFV3	Q8CFV3 mus musculi
9	7	17.1	187 2 Q6G6E9	Q6G6E9 staphylococ
10	7	17.1	209 2 Q6DUT5	Q6DUT5 xenopus lae
11	7	17.1	280 1 ISPE_CLOTE	Q8992a streptidun
12	7	17.1	330 2 Q9XALO	Q9XALO streptomyc
13	7	17.1	334 2 Q8TSU2	Q8TSU2 methanocarb
14	7	17.1	373 2 Q6ING9	Q6ING9 xenopus lae
15	7	17.1	373 2 AAH72312	AAH72312 xenopus l
16	7	17.1	380 2 Q3U037	Q3U037 archaeoglob
17	7	17.1	435 2 Q9WVX3	Q9WVX3 thermotoga
18	7	17.1	467 2 Q9SKC0	Q9SKC0 macaca fasc
19	7	17.1	494 2 Q8KBM0	Q8KBM0 Chlorobium
20	7	17.1	508 2 Q6LOF3	Q6LOF3 picrophilus
21	7	17.1	548 2 Q6DIX6	Q6DIX6 xenopus tro
22	7	17.1	594 2 Q6ZK26	Q6ZK26 homo sapien
23	7	17.1	662 1 BAD18549	BAD18549 homo sap
24	7	17.1	662 1 TLPB_BACSU	P39217 bacillus s
25	7	17.1	681 2 Q6GIX4	Q6GIX4 streptomyc
26	7	17.1	770 2 Q6BM10	Q6BM10 debaryomyc
27	7	17.1	901 2 Q8RSU5	Q8RSU5 thermococ
28	7	17.1	915 2 Q7SG97	Q7SG97 oryza sativ
29	7	17.1	915 2 AA98496	AA98496 oryza sat
30	7	17.1	921 2 Q821Y1	Q821Y1 chlamydomo
31	7	17.1	947 2 Q9LW4	Q9LW4 arabidopsis

32	7	17.1	1002 2 Q9LM94	Q9LM94 arabidopsis
33	7	17.1	1166 2 Q8VSB0	Q8VSB0 shigella fl
34	7	17.1	1209 2 Q72570	Q72570 homo sapien
35	7	17.1	1373 2 Q93QJ7	Q93QJ7 escherichia
36	7	17.1	1623 2 Q9AFU0	Q9AFU0 shigella fl
37	7	17.1	1756 1 TR11_ECOLI	TR11_ECOLI
38	7	17.1	1756 1 TR12_ECOLI	TR12_ECOLI
39	7	17.1	1756 2 Q6S1Z4	Q6S1Z4 escherichia
40	7	17.1	1756 2 Q6TDU5	Q6TDU5 escherichia
41	7	17.1	1756 2 Q84A21	Q84A21 escherichia
42	7	17.1	1756 2 Q9WTB0	Q9WTB0 plasmid r10
43	7	17.1	1756 2 AAQ98619	AAQ98619 escherich
44	7	17.1	1756 2 AAQ25113	AAQ25113 escherich
45	7	17.1	3092 2 Q73XH7	Q73XH7 mycobacteri

ALIGNMENTS

RESULT 1
BP28_HUMAN STANDARD; PRT: 2144 AA.
AC Q9H583; Q9NM23;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Protein BAP28.
GN Name=BAP28;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS SER-1694; ALA-1654; ASP-1667 AND
RP GLY-2017.
RA Bouguieret L., Chumakov I., Barry C., Cohen-Akenine A.;
RT "A novel BAP28 gene and protein."
RL Patent number WO0100669, 04-JAN-2001.
RN [2]
RP SEQUENCE OF 1534-2144 FROM N.A.
RA Cobley V.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1777-2144 FROM N.A.
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Ohtsuka M., Nishitani T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA Nagahara K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA Shiratori A., Sudo H., Hosoi T., Kaku Y., Kodaira H., Kondo H.,
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA Yamazaki M., Ninomiya K., Iishihara T., Yamashita H., Murakawa K.,
RA Fujimori Y., Tanai H., Kimata M., Watanabe M., Hiraoaka S., Chiba Y.,
RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yoshida S., Houchi T.,
RA Kusano J., Kanehori K., Takahashi-Fuji A., Hara H., Tanase T.-O.,
RA Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,
RA Imose N., Masehino K., Yuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
RA Moriya S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Taashiro H., Tanigami A., Fujisawa T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hixio M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;

```

RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs."
RT Nat. Genet. 36:40-45(2004).
CC -1- SIMILARITY: Belongs to the BAP28 family.
CC -1- SIMILARITY: Contains 1 HEAT repeat.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1ab-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AK067150; CAC26776.1; -
CC EMBL; AK067150; CAC26776.1; -
CC EMBL; AK001221; BAA91564.1; ALT_INIT.
CC SWISS-2DPAGE; Q9H583; HUMAN.
CC InterPro; IPR008938; ARM.
CC InterPro; IPR000357; HEAT.
CC PROSITE; PS50077; HEAT_REPEAT; FALSE_NEG.
CC Polymorphism.
CC REPEAT 2106 2142 HEAT.
CC VARIANT 1694 1694 N -> S.
CC VARIANT 1854 1854 V -> A.
CC VARIANT 1967 1967 N -> D.
CC VARIANT 2017 2017 E -> G.
CC SEQUENCE 2144 AA; 242355 MW; D66816E780BC9B7 CRC64;
SO
Query Match 100.0%; Score 41; DB 1; Length 2144;
Best Local Similarity 100.0%; Pred. No. 8,7e-34;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LKLLCKNFGAENPDFFVPLVAVLIPERKEKNVIGSA 41
DB 1674 LKLLCKNFGAENPDFFVPLVAVLIPERKEKNVIGSA 1714
RESULT 2
Q8N7L7 PRELIMINARY; PRT; 897 AA.
AC Q8N7L7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ40893.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=uterus;
Pubmed=14702039;
Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
Sekine M., Ohyashi M., Nishi T., Shibahara T., Tanaka K., Ishii S.,
Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
Tanai H., Kimura M., Watanabe M., Hirano K., Chiba Y., Ishida S.,
Oono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
Kanepori K., Takahashi F., Hara R., Takeuchi K., Arita M., Imose N.,
Miyashiro K., Yumoto H., Oshima A., Sasaki N., Aotsuka S.,
Yoshihara K., Matsunawa H., Ichihara T., Shiohara N., Sano S.,

```

```

RA Moriya S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Taniguchi A., Fujitani Y.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Okawata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Kikawa R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsunuma K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T., Nakagawa K.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Nakamura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs."
RT Nat. Genet. 36:40-45(2004).
CC EMBL; AK098212; BAC05261.1; -
CC InterPro; IPR008938; ARM.
CC SEQUENCE 897 AA; 101574 MW; 5FF6A94FE8855895 CRC64;
SO
Query Match 48.8%; Score 20; DB 2; Length 897;
Best Local Similarity 100.0%; Pred. No. 4.2e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LKLLCKNFGAENPDFFVPLV 20
DB 427 LKLLCKNFGAENPDFFVPLV 446
RESULT 3
ID BP28 MACPA STANDARD; PRT; 958 AA.
AC Q9GM44;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Protein BAP28 (Qmpa-17571) (Fragment).
OS Name=BAP28;
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
OC NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
Otsuda N., Hida M., Kuwada J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries."
RT Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the BAP28 family.
CC -1- SIMILARITY: Contains 1 HEAT repeat.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1ab-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AB049842; BAB16728.1; ALT_INIT.
CC InterPro; IPR008938; ARM.
CC InterPro; IPR000357; HEAT.
CC PROSITE; PS50077; HEAT_REPEAT; FALSE_NEG.
CC NON TER 1
CC REPEAT 920 956 HEAT.
CC SEQUENCE 958 AA; 108644 MW; 3DBD95C3623CFB31 CRC64;
SO
Query Match 48.8%; Score 20; DB 1; Length 958;
Best Local Similarity 100.0%; Pred. No. 4.4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 LKLCXNFGAENPDPFVPL 20
 |||||
 DB 487 LKLCXNFGAENPDPFVPL 506

RESULT 4

06P197 PRELIMINARY; PRT; 1106 AA.
 ID 06P197
 AC 06P197
 DT 05-JUL-2004 (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DE FL10359 protein (Fragment).
 GN Name=FL10359.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshlyuk S., Carinci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallus D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Maita M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RA Strausberg R.L.,
 RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC065205; AAH65205.1; -.
 FT NON TER 1
 SQ SEQUENCE 1106 AA; 125359 MW; 09F7CE94042302C4 CRC64;
 Query Match 48.8%; Score 20; DB 2; Length 1106;
 Best Local Similarity 100.0%; Pred. No. 4.9e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 LKLCXNFGAENPDPFVPL 20
 |||||
 DB 636 LKLCXNFGAENPDPFVPL 655

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshlyuk S., Carinci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallus D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Maita M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RA Strausberg R.L.,
 RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC065205; AAH65205.1; -.
 FT NON TER 1
 SQ SEQUENCE 1106 AA; 125359 MW; 09F7CE94042302C4 CRC64;
 Query Match 48.8%; Score 20; DB 2; Length 1106;
 Best Local Similarity 100.0%; Pred. No. 4.9e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 LKLCXNFGAENPDPFVPL 20
 |||||
 DB 636 LKLCXNFGAENPDPFVPL 655

RESULT 6

0872N5 PRELIMINARY; PRT; 181 AA.
 ID 0872N5
 AC 0872N5
 DT 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 GN Lysozyme, putative.
 OS Pseudomonas syringae (pv. tomato).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=323;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DC3000;
 RX MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;
 RA Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,
 RA Gwin M.L., Dodson R.J., Debey R.T., Durkin A.S., Kolonay J.F.,
 RA Madupu R., Daugherty S.C., Brinkac L.M., Beanan M.J., Haft D.H.,
 RA Nelson W.C., Daviden T.M., Zafar N., Zhou L., Liu J., Yuan Q.,
 RA Khouli H.M., Fedorova N.B., Tran B., Russell D., Berry K.J.,
 RA Utecherback T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,
 RA Deng W.-L., Ramos A.R., Alfano J.R., Cartinhour S., Chatterjee A.K.,
 RA Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,
 RA Bender C.L., White O., Fraser C.M., Collier A.,
 RT "The complete genome sequence of the Arabidopsis and tomato pathogen
 RT Pseudomonas syringae pv. tomato DC3000."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).

```

DR EMBL; AE016868; AA056867.1; -
DR HSP; P23851; ICNS.
DR TIGR; P52703.89; -
DR GO; GO:0004568; F:chitinase activity; IEA.
DR GO; GO:0016998; P:cell wall catabolism; IEA.
DR GO; GO:0006032; P:chitin catabolism; IEA.
DR GO; GO:0009613; P:response to pest/pathogen/parasite; IEA.
DR InterPro; IPR000726; Glyco_hydro_19.
DR Pfam; PF00182; Glyco_hydro_19; 1.
DR ProDom; PD354900; Glyco_hydro_19; 1.
DR Complete proteome.
SQ SEQUENCE 181 AA; 20170 MW; 68F88AD3490A3394 CRC64;

Query Match      19.5%; Score 8; DB 2; Length 181;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      16 FVPVPLNTA 23
      |||||
Db      22 FVPVPLNTA 29

RESULT 7
OGBHY7  PRELIMINARY; PRT; 585 AA.
ID OGBHY7
AC OGBHY7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus adult male hippocampus cDNA, RIKEN full-length enriched
DE library, clone:CG30007C17 product:weakly similar to HYPOTHETICAL 52.2
DE kda PROTEIN (Fragment).
DE Name=CG30007C17Rik;
DE Mus musculus (Mouse);
DE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hippocampus;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hippocampus;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hippocampus;
RA The FANTOM Consortium;
RT the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hippocampus;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hippocampus;
RX MEDLINE=20530913; PubMed=1076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,

```

```

RA Kono H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multichannel sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hippocampus;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katon H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ono M., Ohnishi N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Yamamatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK049881; BAC33969.1; -
DR MGD; MGI:2442949; C630007C17Rik.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; ZnF_C2H2.
DR Pfam; PF00096; Zf-C2H2; 1.
DR SMART; SM00355; ZnF_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR Hypothetical protein.
KM NON TPR 585
FT SEQUENCE 585 AA; 66594 MW; 569ECD93979827A5 CRC64;

Query Match      19.5%; Score 8; DB 2; Length 585;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      14 DPFVPLVN 21
      |||||
Db      418 DPFVPLVN 425

RESULT 8
OGBFY3  PRELIMINARY; PRT; 920 AA.
ID OGBFY3
AC OGBFY3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE C630007C17Rik protein (Fragment).
DE Name=C630007C17Rik;
DE Mus musculus (Mouse);
DE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RX MEDLINE=22388257; PubMed=12477932;
RA Straube R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner C.M., Schuler G.D.,
RA Alschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang J., Hsieh F.,
RA Dichtenko L., Marusik A., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Baha S.S., Loguanello N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Borak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,

```

```

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Smith J., Myers R.W., Butterfield Y.S.,
RA Krzywnski M.I., Skalske U., Smaltus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Maira M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [12]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Straubeberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC055535; AAH55535.1;-.
DR MGI, MGI:2442949; C630007C17RLK.
FT NON_TER
SQ SEQUENCE 920 AA; 103331 MW; 1C8DA479216BA70C CRC64;

Query Match 19.5%; Score 8; DB 2; Length 920;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 DFVPLN 21
Db 138 DFVPLN 145

RESULT 9
AC OGE69 PRELIMINARY; PRT; 187 AA.
ID OGE69
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Putative membrane protein.
GN ORFNames=SR2451;
OS Staphylococcus aureus subsp. aureus MRSA252.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
NCBI_TaxID=282458;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MRSA252;
RA Holden M.T.G., Peil E.J., Lindsay J.A., Peacock S.J., Day N.P.J.,
RA Enright M.C., Foster T.J., Moore C.E., Hurst L., Atkin R., Barton A.,
RA Baeson N., Bentley S.D., Chillingworth C., Chillingworth T.,
RA Churcher C., Clark L., Corcoran C., Cronin A., Doggett J., Dowd L.,
RA Feltwell T., Hance Z., Harris B., Hauser H., Holtroyd S., Jagels K.,
RA James K.D., Leonard N., Line A., Mayes R., Moule S., Mungall K.,
RA Ormond D., Quail M.A., Rabbintovitch E., Rutherford K., Sanders M.,
RA Sharp S., Simmonds M., Stevens K., Whitehead S., Barrett B.G.,
RA Spratt B.G., Parkhill J.;
RT "Complete genomes of two clinical Staphylococcus aureus strains:
RT evidence for the rapid evolution of virulence and drug resistance."
RL Proc. Natl. Acad. Sci. U.S.A. 101:9786-9791(2004).
DR EMBL, BX571856; CNG41433.1;-.
DR InterPro, IPR009057; Homeodomain_1like.
DR InterPro, IPR001647; HTH_Tetr.
SQ SEQUENCE 187 AA; 22098 MW; EEE86B56A000D25 CRC64;

Query Match 17.1%; Score 7; DB 2; Length 187;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 FVPLN 22
Db 130 FVPLN 136

RESULT 10
AC O6DJ5 PRELIMINARY; PRT; 209 AA.
ID O6DJ5

```

```

AC O6DJ5;
DT 01-OCT-2004 (TRENBLrel. 28, Created)
DT 01-OCT-2004 (TRENBLrel. 28, Last sequence update)
DT 01-OCT-2004 (TRENBLrel. 28, Last annotation update)
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodidae; Xenopus.
NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Adamson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huily S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Smith J., Myers R.W., Butterfield Y.S.,
RA Krzywnski M.I., Skalske U., Smaltus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Maira M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [12]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Klein S.L., Straubeberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative."
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Klein S., Straubeberg R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC075194; AAH75194.1;-.
KW Hypothetical protein.
SQ SEQUENCE 209 AA; 23557 MW; B52BF04AB345B39 CRC64;

Query Match 17.1%; Score 7; DB 2; Length 209;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KNEGAE 12
Db 156 KNEGAE 162

RESULT 11
AC O899A2;
ID ISPE CLOTE STANDARD; PRT; 280 AA.
DT 01-OCT-2004 (Rel. 45, Created)
DT 01-OCT-2004 (Rel. 45, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)
DE (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).
GN Name=ISPE; OrderedLocustNames=CTC00283;
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

```

```

OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Massachusetts / E88;
RX MEDLINE=22457253; PubMed=12552129; DOI=10.1073/pnas.0335853100;
RA Bruggemann H., Baumer S., Fricke W.F., Wietzer A., Liesegang H.,
RA Decker I., Heideberg C., Martinez-Arias R., Merti R., Henne A.,
RA Gottschalk G.;
RT "The genome sequence of Clostridium tetani, the causative agent of
RT tetanus disease.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
CC -1- FUNCTION: Catalyzes the phosphorylation of the position 2 hydroxy
CC group of 4-diphosphocytidylyl-2C-methyl-D-erythritol (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-
CC D-erythritol.
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth
CC step.
CC -1- SIMILARITY: Belongs to the GHMP kinase family. IspE subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR HAMM: MF 00061; -; 1.
CC DR EMBL: AE015936; AA034927.1; -.
CC DR InterPro: IPR006204; GHMP_kinase.
CC DR InterPro: IPR000870; Homoser_kin.
CC DR InterPro: IPR004424; IspE.
CC DR Pfam: PF00288; GHMP_kinases; 1.
CC DR PRINTS: PRO0958; HOMSERKINASE.
CC DR TIGRFAMs: TIGR00154; IspE; 1.
CC KM ATP-binding; Complete proteome; Isoprene biosynthesis; Kinase;
CC KM Translasease.
CC KM NP_BIND 91 101 ATP (Potential).
CC FT ACT_SITE 8 8 By similarity.
CC FT ACT_SITE 133 133 By similarity.
CC SQ SEQUENCE 280 AA; 31616 MW; B132AC792B8CDD7F CRC64;

Query Match 17.1%; Score 7; DB 1; Length 280;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LKLLCKN 7
Db 204 LKLLCKN 210

RESULT 12
O9XAL0 PRELIMINARY; PRT; 330 AA.
AC O9XAL0;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative membrane protein.
GN OrderedLocustNames=SC03592; ORFNames=SC6673.03;
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinomycetales; Actinomycetaceae;
OC Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=2196410; PubMed=1200953; DOI=10.1038/41741a;
RX Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RX Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleeer H.,

```

```

RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L.D., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K.M., Ruter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL339117; CAB45459.1; -.
DR PIR: T35360; T35360.
DR InterPro: IPR005135; Exo_endo_phos.
DR Pfam: PF03372; Exo_endo_phos; 1.
KM Complete proteome.
SQ SEQUENCE 330 AA; 34963 MW; 94CDB924DF1B740F CRC64;

Query Match 17.1%; Score 7; DB 2; Length 330;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GAENPDP 15
Db 109 GAENPDP 115

RESULT 13
O8TSU2 PRELIMINARY; PRT; 334 AA.
AC O8TSU2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein MA0703.
GN OrderedLocustNames=MA0703.
OS Methanobacillus acetivorans.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanobacteriaceae; Methanobacteriales.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA FitzHugh W., Calvo S., Engels R., Smitrov S., Atencio D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Graham D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanobacillus acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL: AB010731; AA004143.1; -.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0003887; F:DNA-directed DNA polymerase activity; IEA.
DR GO: GO:0006260; F:DNA replication; IEA.
DR InterPro: IPR003141; Peptidase_PHP_N.
DR Pfam: PF02231; PHP_N; 1.
DR SMART: SM00481; POLIINC; 1.
KM Complete proteome; Hypothetical protein.
SQ SEQUENCE 334 AA; 38654 MW; 1BCACBA84752B316 CRC64;

Query Match 17.1%; Score 7; DB 2; Length 334;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KLIPER 31

```

Db 11 KLIAPER 17

RESULT 14
 Q6ING9 PRELIMINARY; PRT; 373 AA.
 ID Q6ING9
 AC Q6ING9
 DT 05-JUL-2004 (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE MG082621 protein.
 GN Name=MG082621;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 OC Xenopodinae; Xenopus.
 NC NCB1_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Boulford G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallov D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=22341132; PubMed=12454917;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.;
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 RT Initiative.";
 RL Dev. Dyn. 225:384-391(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX Klein S., Strausberg R.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: Belongs to the PP2C family.
 DR EMBL; BC072312; AAH72312.1; -
 DR InterPro; IPR000222; PP2C.
 DR InterPro; IPR001932; PP2C-like.
 DR Pfam; PF00481; PP2C; 1.
 DR SMART; SM00332; PP2CC; 1.
 DR PROSITE; PS01032; PP2C; 1.
 KW Hydroxylase; Magnesium.
 SQ SEQUENCE 373 AA; 41536 MW; 58713D1A52099BB7 CRC64;

Query Match 17.1%; Score 7; DB 2; Length 373;
 Best Local Similarity 100.0%; Pred. No. 83;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 PERKEEK 35
 Db 229 PERKEEK 235

RESULT 15
 AAH72312 PRELIMINARY; PRT; 373 AA.
 ID AAH72312
 AC AAH72312
 DT 01-JUN-2004 (TREMBlrel. 27, Created)
 DT 01-JUN-2004 (TREMBlrel. 27, Last sequence update)
 DT 01-JUN-2004 (TREMBlrel. 27, Last annotation update)
 DE Hypothetical protein.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 OC Xenopodinae; Xenopus.
 NC NCB1_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=22341132; PubMed=12454917;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.;
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 RT Initiative.";
 RL Dev. Dyn. 225:384-391(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Boulford G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallov D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX Klein S., Strausberg R.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC072312; AAH72312.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 373 AA; 41536 MW; 58713D1A52099BB7 CRC64;

Query Match 17.1%; Score 7; DB 2; Length 373;
 Best Local Similarity 100.0%; Pred. No. 83;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 PERKEEK 35
 Db 229 PERKEEK 235

Search completed: February 1, 2005, 15:30:59
 Job time : 81.5556 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:53:01 / Search time 69.7778 Seconds
(without alignments)
210.782 Million cell updates/sec

Title: SEQ5ASN1967

Perfect score: 41
Sequence: 1 LKGLFTLFAHLVLPFADTL.....QVNISKDEAFDSENDER 41

Scoring table: OLIGO

Gapop 60.0, Gapext 60.0

Searched: 2002273 seqs, 358729299 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1316031

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: A_Geneseq_23Sep04.*

1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Match Length	ID	Description
1	41	100.0	349 4	AAB92729 Human pro
2	41	100.0	1149 7	ADE08012 Novel pro
3	20	48.8	515 2	AAM54099 Homo sapi
4	20	48.8	2144 4	AAB85029 Protein e
5	7	17.1	220 2	AAY23793 A gyrase
6	7	17.1	453 5	ABBS5123 Lactococ
7	7	17.1	1045 2	ABP60399 Yeast HMG
8	7	17.1	1045 2	AARS611 Yeast HMG
9	7	17.1	1045 8	ADM98779 HMG-COA r
10	7	17.1	1045 8	ADM98798 HMG-COA r
11	7	17.1	1045 8	ADM98844 HMG-COA r
12	7	17.1	1045 8	ADM98916 HMG-COA r
13	7	17.1	1045 8	ADM98889 HMG-COA r
14	6	14.6	51 4	AAM21705 Peptide #
15	6	14.6	51 4	ABBA4072 Peptide #
16	6	14.6	55 7	ADC01241 Enterococ
17	6	14.6	65 7	ADT07598 Bacteri
18	6	14.6	71 5	ABP03187 Human ORF
19	6	14.6	74 5	ABP02661 Human ORF
20	6	14.6	85 4	ABG02711 Novel hum
21	6	14.6	112 3	ABBA1436 Human ORF
22	6	14.6	112 3	ABP00012 Human ORF
23	6	14.6	115 8	ADM98791 HMG-COA r
24	6	14.6	115 8	ADM98931 HMG-COA r
25	6	14.6	118 4	AAM51702 Human tub

26	6	14.6	120 4	AA81563 S. epider
27	6	14.6	120 6	ABM71043 Staphyloc
28	6	14.6	121 5	ABP38311 Staphyloc
29	6	14.6	129 3	AAV52997 Neisseria
30	6	14.6	133 4	ABG22899 Novel hum
31	6	14.6	160 6	ADA35306 Acinetoba
32	6	14.6	185 6	ABU01034 S. pneumo
33	6	14.6	187 8	ADK48744 Streptoco
34	6	14.6	199 3	AAV59939 Arabidops
35	6	14.6	199 3	AAV59937 Arabidops
36	6	14.6	200 2	AAV22570 Bacteri
37	6	14.6	201 3	AAV59938 Arabidops
38	6	14.6	201 3	AAV59935 Arabidops
39	6	14.6	201 8	ADN74077 Thale cre
40	6	14.6	208 3	AAV59937 Arabidops
41	6	14.6	208 3	AAV59937 Arabidops
42	6	14.6	214 2	AAV48812 Human KDE
43	6	14.6	214 5	ABP64856 Human pro
44	6	14.6	214 8	ADQ19937 Human sof
45	6	14.6	217 7	ABM74402 DNA clone

ALIGNMENTS

RESULT 1
AAB92729
ID AAB92729 standard; protein: 349 AA.
XX
AC AAB92729;
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:11159.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EPI074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX
PS Claim 8; SEQ ID NO 11159; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH13742 represent human cDNA sequences; AAH92446 to AAH95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention

CC Sequence 349 AA;

Query Match 100.0%; Score 41; DB 4; Length 349;
 Best Local Similarity 100.0%; Pred. No. 4.5e-36;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LKGLFTLPAGHLVKKFPADTLNQVNISKTDPAFPDSENDPEK 41
 DB 152 LKGLFTLPAGHLVKKFPADTLNQVNISKTDPAFPDSENDPEK 192

RESULT 2

ADDE08012
 ID ADE08012 standard; protein; 1149 AA.

AC ADE08012;

DT 29-JAN-2004 (first entry)

DE Novel protein (useful for identifying genetic disorders) #167.

XX novel gene; novel protein; tissue marker; molecular weight marker;
 KW chromosome marker; genetic disorder.

XX Unidentified.

PN WO2003054152-A2.

XX 03-JUL-2003.

PF 10-DEC-2002; 2002WO-US039555.

XX 10-DEC-2001; 2001US-0339739P.

PR 11-DEC-2001; 2001US-0339453P.

PR 14-MAR-2002; 2002US-0365091P.

PR 14-MAR-2002; 2002US-0365384P.

PR 12-APR-2002; 2002US-0372381P.

PR 12-APR-2002; 2002US-0372615P.

PR 22-APR-2002; 2002US-00128558.

PR 24-APR-2002; 2002US-0376045P.

XX (HYSE-) HYSEQ INC.

PA WPI: 2003-569235/53.

DR N-PSDB; ADE07101.

XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
 PI Ghosh M, Xue AV, Wehtman T, Weng G, Zhou P, Drmanac RT, Wang Z;
 PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;

XX WPI: 2003-569235/53.

CC New polynucleotides, useful for expressing recombinant proteins for
 PT analysis, characterization or therapeutic use, or as markers for tissues
 PT in which the corresponding protein is preferentially expressed.

XX Claim 20; SEQ ID NO 1078; 1177pp; English.

CC The invention comprises the amino acid and coding sequences of novel
 CC proteins. The DNA and protein sequences of the invention are useful as:

CC markers for tissues in which the corresponding protein is preferentially
 CC expressed; as molecular weight markers on gels; as chromosome markers or
 CC tags; to identify chromosomes or to map related gene positions; and to
 CC compare with endogenous DNA sequences in patients to identify potential
 CC genetic disorders. The present amino acid sequence represents a protein
 CC of the invention.

SO Sequence 1149 AA;

Query Match 100.0%; Score 41; DB 7; Length 1149;
 Best Local Similarity 100.0%; Pred. No. 1.3e-35;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LKGLFTLPAGHLVKKFPADTLNQVNISKTDPAFPDSENDPEK 41
 DB 952 LKGLFTLPAGHLVKKFPADTLNQVNISKTDPAFPDSENDPEK 992

RESULT 3

AAW54099
 ID AAW54099 standard; protein; 515 AA.

AC AAW54099;

DT 28-SEP-1998 (first entry)

DE Homo sapiens BAP28 sequence.

XX BARD1; ring protein; BRCA1; breast cancer; risk; diagnosis.

XX Homo sapiens.

XX WO9812327-A2.

XX 26-MAR-1998.

PD 19-SEP-1997; 97WO-US016842.

PF 20-SEP-1996; 96US-0025296P.

PR 03-APR-1997; 97US-0042611P.

PR 04-APR-1997; 97US-0042985P.

XX (TEXA) UNIV TEXAS SYSTEM.

XX Bowcock AM, Baer R;

XX WPI: 1998-230317/20.

DR N-PSDB; AAW24135.

XX DNA sequence encoding BARD1, B123, BE2, BE14, BE31 or BE445 - which as
 PT breast cancer antigen, BRCA1, binding proteins are useful to identify
 PT patient having or at risk of developing cancer.

XX Disclosure; Page 287-288; 348pp; English.

XX The sequence is that of a protein which can be used in the preparation of
 CC the recombinant breast cancer antigen, BRCA1, binding proteins BARD1,
 CC B123, BE2, BE14, BE31 or BE445, or a composition for the detection of a
 CC BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid sequence, specifically
 CC a wild type BARD1 composition for the detection or purification of BRCA1,
 CC useful to identify a patient having, or at risk of developing cancer.

CC BARD1 can be used in the preparation of an anti-BARD1 antibody, and in
 CC the detection and purification of a BRCA1 protein. BARD1, B123, BE2,
 CC BE14, BE31 or BE445 can be used in the identification of a binding protein
 CC agonist or antagonist that alters the binding of BARD1, B123, BE2, BE14,
 CC BE31 or BE445 to BRCA1 or the biological activity of the BRCA1-BARD1,
 CC B123, BE2, BE14, BE31 or BE445 complex. The antibodies can be used to
 CC detect BARD1, B123, BE2, BE14, BE31 or BE445, a specific anti-BARD1
 CC antibody can be used to identify a patient having or at risk of
 CC developing cancer

XX Sequence 515 AA;

Query Match 48.8%; Score 20; DB 2; Length 515;
 Best Local Similarity 100.0%; Pred. No. 4e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LKGLFTLPAGHLVKPFADTL 20
 |||||
 318 LKGLFTLPAGHLVKPFADTL 337

RESULT 4

AAB85029
 ID AAB85029 standard; protein; 2144 AA.

XX AAB85029;

DT 06-AUG-2001 (first entry)

XX Protein encoded by BAP28 cDNA consisting of exons 1 to 45.

XX BAP28; prostate; tumour; cancer; diagnostic; genetic analysis.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 1694 /label= Ser or Asn

FT Misc-difference 1854 /label= Ala or Val

FT Misc-difference 1967 /label= Asp or Asn

FT Misc-difference 2017 /label= Gly or Glu

XX W0200100669-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000MO-IB001183.

XX 25-JUN-1999; 99US-0141323P.

XX 18-JAN-2000; 2000US-0176880P.

XX (GEST) GENSET.

XX Barry C, Bougueleret L, Chumakov I, Cohen-Akenine A;

XX WPI; 2001-367032/38.

XX N-PSDB; AAF83909, AAF83910.

XX New BAP28 polynucleotides and polypeptides overexpressed in prostate

XX cancer cells for diagnosing prostate tumour, e.g. by hybridization or

XX polymerase chain reaction assays.

XX Claim 14; Page 297-304; 349pp; English.

XX The invention is directed to BAP28 polypeptides, BAP28 polynucleotide
 CC sequences and regulatory region located at the 3' and 5' ends of the
 CC BAP28 coding region. The BAP28 polypeptides can be expressed by standard
 CC recombinant methodology. BAP28 polynucleotides and polypeptides have been
 CC found to be over expressed in prostate tumour cells, therefore levels of
 CC BAP28 expression and/or activity may be assayed (e.g. by polymerase chain
 CC reaction (PCR)) to diagnose patient suffering from or susceptible to
 CC prostate cancer. Antibodies specific for the BAP28 polypeptides are
 CC useful as diagnostic reagents. Allelic markers of the BAP28 gene are
 CC useful in genetic analysis. The present sequence represents a protein
 CC encoded by a first cDNA sequence of the BAP28 gene consisting of the
 CC exons 1 to 45

XX Sequence 2144 AA;

Query Match 48.8%; Score 20; DB 4; Length 2144;
 Best Local Similarity 100.0%; Pred. No. 1.4e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LKGLFTLPAGHLVKPFADTL 20
 |||||
 Db 1947 LKGLFTLPAGHLVKPFADTL 1966

RESULT 5

AAY23793
 ID AAY23793 standard; protein; 220 AA.

XX AAY23793;

XX 27-AUG-2003 (revised)

DT 14-SEP-1999 (first entry)

XX A gyrase protein of Chitinophaga pinensis.

XX Identification; detection; microbe; gyrase gene; gyrase protein.

XX Chitinophaga pinensis.

XX JP11169175-A.

XX 29-JUN-1999.

XX 12-DEC-1997; 97JP-00343316.

XX 12-DEC-1997; 97JP-00343316.

XX (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.

XX WPI; 1999-422615/36.

XX N-PSDB; AAX86000.

XX Identification and detection of a microbe - by detection of a gyrase

XX gene.

XX Example 3; Page 14-15; 42pp; Japanese.

XX The specification describes a method for the identification or detection
 CC of a microbe, using the gyrase gene as the index. The method involves the
 CC use of PCR primers to amplify DNA from the microbe, which is then
 CC identified or detected depending on its base sequence. The method can be
 CC used to classify and identify an unidentified microbe strain rapidly and
 CC with high precision. The present sequence represents a gyrase protein.
 CC (Updated on 27-AUG-2003 to correct OS field.)

XX Sequence 220 AA;

Query Match 17.1%; Score 7; DB 2; Length 220;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 SENDPER 41
 |||||
 Db 130 SENDPER 136

RESULT 6
 ID ABB55123 standard; protein; 453 AA.

XX ABB55123;

XX 29-AUG-2003 (revised)

DT 16-MAY-2002 (first entry)

XX Lactococcus lactis protein yeda.

XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.

XX Lactococcus lactis; IL1403.

PN FR2807446-A1.
 XX 12-OCT-2001.
 XX 11-APR-2000; 2000FR-00004630.
 XX 11-APR-2000; 2000FR-00004630.
 XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
 XX Bolocrine A, Sorokine A, Renault P, Ehrlich SD;
 XX WPI; 2002-043418/06.
 XX New nucleotide sequence useful in the identification or Lactococcus
 XX lactis and related species.
 XX Claim 6; SEQ ID NO 1825; 2504bp; French.
 XX The present invention is related to a Lactococcus lactis nucleotide
 XX sequence (ABA90521) and related proteins (ABR53300-ABR55621). The nucleic
 XX acid sequence is useful in the detection and/or amplification of nucleic
 XX acid sequence, particularly to identify Lactococcus lactis or related
 XX species. The proteins of the invention are useful for the biosynthesis or
 XX biodegradation of a composition of interest. The invention helps research
 XX in lactic bacteria, particularly useful in the production of yogurt and
 XX cheese. Note: The sequence data for this patent is based on equivalent
 XX patent WO2001/7334 (published 18-OCT-2001) which is available in
 XX electronic format directly from WIPO at
 XX ftp://wipo.int/pub/published_pct_sequences. (Updated on 29-AUG-2003 to
 XX standardise OS field)
 XX Sequence 453 AA;
 SQ
 Query Match 17.1%; Score 7; DB 5; Length 453;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 22 QVNISKT 28
 DB 127 QVNISKT 133
 RESULT 7
 ABB60399
 ID ABB60399 standard; protein; 1045 AA.
 XX ABB60399;
 XX 21-MAR-2003 (first entry)
 XX Yeast HMG2 SEQ ID NO 6.
 XX Yeast; HMG-CoA reductase; squalene; zymosterol; cholesta-7,24-dienol;
 XX cholesta-5,7,24-trienol; zymosterol-24-methyl transferase;
 XX ergosta-5,7,24(28)-trienol-22-dehydrogenase; ergosta; dienol;
 XX episterol-5-dehydrogenase; linker region; catalytic domain;
 XX membrane binding region; HMG2; enzyme.
 XX Saccharomyces cerevisiae.
 XX US5460949-A.
 XX 24-OCT-1995.
 XX 28-OCT-1991; 91US-00783861.
 XX 15-NOV-1990; 90US-00613380.
 XX (STAD) AMOCO CORP.
 XX Mukharji I, Saunders CA, Wolf FR;
 XX

DR WPI; 1992-168867/21.
 DR N-PSDB; ABZ26037.
 XX Increasing squalene and specific sterol accumulation in yeasts - by
 XX transforming mutant yeasts to increase 3-hydroxy-3-methyl:glutaryl CoA
 XX reductase activity in the yeasts.
 XX Disclosure; Col 63-68; 60pp; English.
 XX The invention relates to: (A) a method of increasing squalene,
 XX zymosterol, cholesta-7,24-dienol and cholesta-5,7,24-trienol
 XX accumulation in yeast comprising increasing the expression level of a
 XX structural gene encoding a polypeptide having HMG-CoA reductase activity
 XX in a mutant yeast having defects in the expression of zymosterol-24-
 XX methyl transferase and ergosta-5,7,24(28)-trienol-22-dehydrogenase;
 XX (B) a method of increasing squalene, ergosta-8,22-dienol, ergosta-7,22-
 XX dienol, ergosta-8-enol and ergosta-7-enol accumulation in S. cerevisiae
 XX comprising transforming a mutant S. cerevisiae having a defect in the
 XX expression of episterol-5-dehydrogenase with a recombinant DNA molecule
 XX comprising a vector operatively linked to an exogenous DNA segment that
 XX encodes the catalytic region and at least a portion of the linker region
 XX but is free from the membrane binding region of an HMG-CoA reductase
 XX enzyme and a promoter suitable for driving the expression of the
 XX reductase in the yeast; (C) a method of increasing squalene, zymosterol
 XX and cholesta-7,24-dienol accumulation in S. cerevisiae comprising
 XX transforming a mutant S. cerevisiae having a defect in the expression of
 XX zymosterol-24-methyl transferase and episterol-5-dehydrogenase with a
 XX recombinant DNA molecule as in (B); (D) a method of increasing squalene,
 XX zymosterol, ergosta-5,7,24(28)-trienol and ergosta-5,7-dienol
 XX accumulation in S. cerevisiae comprising transforming a mutant S.
 XX cerevisiae having a defect in the expression of ergosta-5,7,24(28)-
 XX trienol-22-dehydrogenase with a recombinant DNA molecule as in (B); (E) a
 XX mutant S. cerevisiae having defects in the expression of zymosterol-24-
 XX methyl transferase and ergosta-5,7,24(28)-trienol-22-dehydrogenase
 XX enzymes, which mutant is designated ATC0402mm; (F) a mutant of S.
 XX cerevisiae having single or double defects in the expression of enzymes
 XX that catalyse the conversion of squalene to ergosterol, transformed with
 XX a recombinant DNA molecule as in (B). The present sequence is that of the
 XX Saccharomyces cerevisiae HMG-CoA reductase 2 (HGM2) protein of the
 XX invention
 SQ Sequence 1045 AA;
 Query Match 17.1%; Score 7; DB 2; Length 1045;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 HLYKPPA 17
 DB 10 HLYKPPA 16
 RESULT 8
 AAR58611
 ID AAR58611 standard; protein; 1045 AA.
 XX AAR58611;
 XX 25-MAR-2003 (revised)
 XX 28-APR-1995 (first entry)
 XX Yeast HMG-CoA reductase 2.
 XX HMG-CoA reductase 2; 3-hydroxy-3-methyl:glutaryl coenzyme A; squalene;
 XX sterol.
 XX Saccharomyces cerevisiae.
 XX US5349126-A.
 XX 20-SEP-1994.
 XX 14-AUG-1992; 92US-00934374.
 XX

```
XX 12-OCT-1990; 90US-00596467.
XX (STAD ) AMOCO CORP.
XX
XX Wolf FR, Saunders CA, Chappell J;
XX WPI; 1994-302280/37.
XX N-PSDB; AAQ70611.
XX
XX Transgenic plants with increased insect resistance - by transformation
XX with DNA encoding HMG-CoA activity to increase squalene and sterol
XX accumulation.
XX
XX Disclosure; Page 53; 58pp; English.
XX
XX AAQ70611 codes for a yeast HMG-CoA reductase 2, AAR56611. Using the A.
XX tumefaciens vector pKYLX71, AAQ70611 was used to produce transgenic
XX plants with increased HMG-CoA reductase activity. Plants with increased
XX HMG-CoA reductase activity were found to have greater squalene and sterol
XX accumulation, and therefore increased insect resistance. (Updated on 25-
XX MAR-2003 to correct PF field.)
XX
XX Sequence 1045 AA;
SQ
Query Match 17.1%; Score 7; DB 2; Length 1045;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 HLVPKFA 17
Db 10 HLVPKFA 16
RESULT 9
ADM98779
ID ADM98779 standard; protein; 1045 AA.
XX
XX ADM98779;
AC
XX 01-JUL-2004 (first entry)
DT
XX
XX HMG-CoA reductase polypeptide #32.
DE
XX
XX Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;
XX diterpene synthase; defence toxin; volatile defensive signal;
XX pollinator attractant; photoprotectant; HMG-CoA reductase; enzyme.
XX
XX Saccharomyces cerevisiae.
OS
XX US2004072323-A1.
PN
XX 15-APR-2004.
PD
XX
XX 07-JAN-2002; 2002US-00041018.
PF
XX
XX 05-JAN-2001; 2001US-0259880P.
PR
XX
XX (MATS/) MATSUDA S P T.
XX (HART/) HART E A.
XX
XX Matsuda SPT, Hart EA;
XX
XX WPI; 2004-373921/35.
XX
XX New unicellular organisms comprising exogenous nucleic acids encoding a
XX geranylgeranyl pyrophosphate and a diterpene synthase, useful for
XX producing diterpene and diterpene precursors.
XX
XX Disclosure; SEQ ID NO 199; 38pp; English.
XX
XX The invention relates to a unicellular organism for producing a diterpene
XX or diterpene precursor comprising an exogenous nucleic acid sequence
```

```
CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
CC promoter operable in the organism, and an exogenous nucleic acid sequence
CC encoding a diterpene synthase under the control of a promoter operable in
CC the organism. The invention also relates to methods of producing a
CC diterpene or diterpene precursor and a method of isolating a diterpene
CC synthase comprising growing several cells in the presence of a
CC polyaromatic resin to make a cell/resin mixture, where at least one of
CC the cells further comprises at least one isolated and purified nucleic
CC acid sequence of a yeast expression library, and the expression of the
CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents an HMG-CoA reductase polypeptide used in the scope of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification but was obtained in electronic format from
CC USPTO at seqdata.uspto.gov/sequence.html.
XX
XX Sequence 1045 AA;
SQ
Query Match 17.1%; Score 7; DB 8; Length 1045;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 HLVPKFA 17
Db 10 HLVPKFA 16
RESULT 10
ADM98798
ID ADM98798 standard; protein; 1045 AA.
XX
XX ADM98798;
AC
XX 01-JUL-2004 (first entry)
DT
XX
XX HMG-CoA reductase polypeptide #51.
DE
XX
XX Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;
XX diterpene synthase; defence toxin; volatile defensive signal;
XX pollinator attractant; photoprotectant; HMG-CoA reductase; enzyme.
XX
XX Saccharomyces cerevisiae.
OS
XX US2004072323-A1.
PN
XX 15-APR-2004.
PD
XX
XX 07-JAN-2002; 2002US-00041018.
PF
XX
XX 05-JAN-2001; 2001US-0259880P.
PR
XX
XX (MATS/) MATSUDA S P T.
XX (HART/) HART E A.
XX
XX Matsuda SPT, Hart EA;
XX
XX WPI; 2004-373921/35.
XX
XX New unicellular organisms comprising exogenous nucleic acids encoding a
XX geranylgeranyl pyrophosphate and a diterpene synthase, useful for
XX producing diterpene and diterpene precursors.
XX
XX Disclosure; SEQ ID NO 218; 38pp; English.
XX
XX The invention relates to a unicellular organism for producing a diterpene
XX or diterpene precursor comprising an exogenous nucleic acid sequence
```

CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
CC promoter operable in the organism, and an exogenous nucleic acid sequence
CC encoding a diterpene synthase under the control of a promoter operable in
CC the organism. The invention also relates to methods of producing a
CC diterpene or diterpene precursor and a method of isolating a diterpene
CC synthase comprising growing several cells in the presence of a
CC polyaromatic resin to make a cell/resin mixture, where at least one of
CC the cells further comprises at least one isolated and purified nucleic
CC acid sequence of a yeast expression library, and the expression of the
CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents an HMG-CoA reductase polypeptide used in the scope of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification but was obtained in electronic format from
CC USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 1045 AA;

Query Match 17.1%; Score 7; DB 8; Length 1045;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 HLKPPA 17
|||
10 HLKPPA 16

Db

RESULT 11
ADM98844 ID ADM98844 standard; protein; 1045 AA.

XX ADM98844;

XX 01-JUL-2004 (first entry)

XX HMG-CoA reductase polypeptide #97.

XX Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;
XX diterpene synthase; defence toxin; volatile defensive signal;
XX pollinator attractant; photoprotectant; HMG-CoA reductase; enzyme.

XX Saccharomyces cerevisiae.

XX US2004072323-A1.

XX 15-APR-2004.

XX 07-JAN-2002; 2002US-00041018.

XX 05-JAN-2001; 2001US-0259880P.

XX (MATS/) MATSUDA S P T.

XX (HART/) HART E A.

XX Matsuda SPT, Hart EA;

XX WPI; 2004-373921/35.

XX New unicellular organisms comprising exogenous nucleic acids encoding a
XX geranylgeranyl pyrophosphate and a diterpene synthase, useful for
XX producing diterpenes and diterpene precursors.

XX Disclosure; SEQ ID NO 264; 38pp; English.

XX The invention relates to a unicellular organism for producing a diterpene
XX or diterpene precursor comprising an exogenous nucleic acid sequence

CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
CC promoter operable in the organism, and an exogenous nucleic acid sequence
CC encoding a diterpene synthase under the control of a promoter operable in
CC the organism. The invention also relates to methods of producing a
CC diterpene or diterpene precursor and a method of isolating a diterpene
CC synthase comprising growing several cells in the presence of a
CC polyaromatic resin to make a cell/resin mixture, where at least one of
CC the cells further comprises at least one isolated and purified nucleic
CC acid sequence of a yeast expression library, and the expression of the
CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents an HMG-CoA reductase polypeptide used in the scope of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification but was obtained in electronic format from
CC USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 1045 AA;

Query Match 17.1%; Score 7; DB 8; Length 1045;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 HLKPPA 17
|||
10 HLKPPA 16

Db

RESULT 12
ADM98916 ID ADM98916 standard; protein; 1045 AA.

XX ADM98916;

XX 01-JUL-2004 (first entry)

XX HMG-CoA reductase polypeptide #169.

XX Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;
XX diterpene synthase; defence toxin; volatile defensive signal;
XX pollinator attractant; photoprotectant; HMG-CoA reductase; enzyme.

XX Saccharomyces cerevisiae.

XX US2004072323-A1.

XX 15-APR-2004.

XX 07-JAN-2002; 2002US-00041018.

XX 05-JAN-2001; 2001US-0259880P.

XX (MATS/) MATSUDA S P T.

XX (HART/) HART E A.

XX Matsuda SPT, Hart EA;

XX WPI; 2004-373921/35.

XX New unicellular organisms comprising exogenous nucleic acids encoding a
XX geranylgeranyl pyrophosphate and a diterpene synthase, useful for
XX producing diterpenes and diterpene precursors.

XX Disclosure; SEQ ID NO 336; 38pp; English.

XX The invention relates to a unicellular organism for producing a diterpene
XX or diterpene precursor comprising an exogenous nucleic acid sequence

CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
CC promoter operable in the organism, and an exogenous nucleic acid sequence
CC encoding a diterpene synthase under the control of a promoter operable in
CC the organism. The invention also relates to methods of producing a
CC diterpene or diterpene precursor and a method of isolating a diterpene
CC synthase comprising growing several cells in the presence of a
CC polyaromatic resin to make a cell/resin mixture, where at least one of
CC the cells further comprises at least one isolated and purified nucleic
CC acid sequence of a yeast expression library, and the expression of the
CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents an HMG-CoA reductase polypeptide used in the scope of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification but was obtained in electronic format from
CC USPTO at seqdata.uspto.gov/sequence.html.
CC
XX
SQ Sequence 1045 AA;

Query Match 17.1%; Score 7; DB 8; Length 1045;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 HLVKPPA 17
|||
Db 10 HLVKPPA 16

RESULT 13
ADM9889
ID ADM9889 standard; protein; 1045 AA.

XX AC ADM9889;

XX DT 01-JUN-2004 (first entry)

XX DE HMG-CoA reductase polypeptide #142.

XX KM Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;
XX KW diterpene synthase; defence toxin; volatile defensive signal;
XX KM pollinator attractant; photoprotectant; HMG-CoA reductase; enzyme.

XX OS Saccharomycetes cerevisiae.

XX PN US2004072323-A1.

XX PN 15-APR-2004.

XX PD 07-JAN-2002; 2002US-00041018.

XX PF 05-JAN-2001; 2001US-0259880P.

XX PR (MATSU) MATSUDA S P T.

XX PA (HART) HART E A.

XX PI Mateuda SPT, Hart EA;

XX DR WPI; 2004-373921/35.

XX PT New unicellular organisms comprising exogenous nucleic acids encoding a
XX geranylgeranyl pyrophosphate and a diterpene synthase, useful for
XX producing diterpenes and diterpene precursors.

XX PS Disclosure; SEQ ID NO 309; 38pp; English.

XX CC The invention relates to a unicellular organism for producing a diterpene
XX or diterpene precursor comprising an exogenous nucleic acid sequence

CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
CC promoter operable in the organism, and an exogenous nucleic acid sequence
CC encoding a diterpene synthase under the control of a promoter operable in
CC the organism. The invention also relates to methods of producing a
CC diterpene or diterpene precursor and a method of isolating a diterpene
CC synthase comprising growing several cells in the presence of a
CC polyaromatic resin to make a cell/resin mixture, where at least one of
CC the cells further comprises at least one isolated and purified nucleic
CC acid sequence of a yeast expression library, and the expression of the
CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents an HMG-CoA reductase polypeptide used in the scope of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification but was obtained in electronic format from
CC USPTO at seqdata.uspto.gov/sequence.html.
CC
XX
SQ Sequence 1045 AA;

Query Match 17.1%; Score 7; DB 8; Length 1045;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 HLVKPPA 17
|||
Db 10 HLVKPPA 16

RESULT 14
AAM21705
ID AAM21705 standard; protein; 51 AA.

XX AC AAM21705;

XX DT 12-OCT-2001 (first entry)

XX DE Peptide #8139 encoded by probe for measuring cervical gene expression.

XX KM Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer.

XX OS Homo sapiens.

XX PN WO200157278-A2.

XX PN 09-AUG-2001.

XX PD 30-JAN-2001; 2001WO-US000670.

XX PF 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488901/53.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human cervical epithelial cells.

XX PS Claim 27; SEQ ID NO 26531; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes
 CC (SENP: see A110068-A128459). The present sequence is a peptide encoded
 CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer. Note: The sequence data for this patent did not form
 CC part of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 CC
 SQ Sequence 51 AA;

Query Match 14.6%; Score 6; DB 4; Length 51;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 AGHLVK 14
 |||||
 DB 40 AGHLVK 45

RESULT 15

ABB44072
 ID ABB44072 standard; peptide; 51 AA.

AC ABB44072;

DT 04-FEB-2002 (first entry)

DE Peptide #11578 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

OS Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000669.

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Claim 27; SEQ ID NO 36707; 639pp + Sequence listing; English.

CC The invention relates to a single exon nucleic acid probe for measuring
 CC human gene expression in a sample derived from human foetal liver. The
 CC single exon nucleic acid probes may be used for predicting, measuring and
 CC displaying gene expression in samples derived from human foetal liver. The
 CC present sequence is a peptide encoded by a single exon nucleic acid probe
 CC of the invention. Note: The sequence data for this patent did not form
 CC part of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 CC
 SQ Sequence 51 AA;

Query Match 14.6%; Score 6; DB 4; Length 51;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AGHLVK 14
 |||||
 DB 40 AGHLVK 45

Search completed: February 1, 2005, 15:20:17
 Job time : 70.7778 secs

16-OCT-2001 (Rel. 40, Created)
 16-OCT-2001 (Rel. 40, Last sequence update)
 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Protein BAP28.
 GN Name=BAP28;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANTS SER-1694; ALA-1854; ASP-1967 AND
 RP GLY-2017.
 RA Bouqueret L., Chumakov I., Barry C., Cohen-Akenine A.;
 RT "A novel BAP28 gene and protein."
 RL Patent number WO0100669, 04-JAN-2001.
 RN [2]
 RP SEQUENCE OF 1534-2144 FROM N.A.
 RA Cobley V.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1777-2144 FROM N.A.
 RA PubMed=14702039; DOI=10.1038/ng1285;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Makamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J., Ii, Saito K., Kawai Y., Isono Y., Nakamura Y.,
 RA Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
 RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
 RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
 RA Omura Y., Abe K., Kamihara K., Katsura N., Sato K., Tanikawa M.,
 RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
 RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hirakawa S., Chiba Y.,
 RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Houchi T.,
 RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
 RA Nomura Y., Mueshino K., Komai F., Hara R., Takeuchi K., Arita M.,
 RA Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotaka S.,
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 RA Hishigaki H., Watanabe K., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirota M., Ohmori Y.,
 RA Kawanabe A., Hikiji T., Kobayashi T., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
 RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu Y., Nakagawa K.,
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuo Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
 RA "Complete sequencing and characterization of 21,243 full-length human
 RT cDNAs.";
 RL Nat. Genet. 36:40-45(2004).
 CC -1 SIMILARITY: Belongs to the BAP28 family.
 CC -1 SIMILARITY: Contains 1 HEAT repeat.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).

FT VARIANT 1694 1694 N -> S
 FT FT /FTID=VAR_010939.
 FT VARIANT 1854 1854 V -> A.
 FT FT /FTID=VAR_010940.
 FT VARIANT 1967 1967 N -> D.
 FT FT /FTID=VAR_010941.
 FT VARIANT 2017 2017 E -> G.
 FT FT /FTID=VAR_010942.
 SO SEQUENCE 2144 AA, 242355 MW, D66816E78D8C9B7 CRC64;
 Query Match 100.0%; Score 41; DB 1; Length 2144;
 Best local Similarity 100.0%; Pred. No. 6,1e-34;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 LKGLFTLPAGHLVKEPFDITLQVNIKSTDAFPDSEDPK 41
 Db 1947 LKGLFTLPAGHLVKEPFDITLQVNIKSTDAFPDSEDPK 1987
 RESULT 3
 ID 08VCX1 PRELIMINARY; PRT; 349 AA.
 AC 08VCX1;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE BC019693 protein.
 GN Name=BC019693;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA PubMed=12388257; PubMed=12477932;
 RA MEDLINE=22388257; PubMed=12477932;
 RA Kraussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Stausberg R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
 RA Diatchenko L., Marisana K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loggiano N.A., Peters G.J., Abramson R.D., Mallory S.J.,
 RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Rahney J., Helton E., Kettelman W., Madan A.C., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywicki M.J., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA STRAIN=mx FVB/N;
 RC STRAIN=mx FVB/N;
 RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
 RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC019693; AAH19693.1;
 DR MGD; MGI:2384963; BC019693.
 DR InterPro; IPR008938; ARM.
 SO SEQUENCE 349 AA, 40166 MW, 9763D0331AD0F515 CRC64;
 Query Match 87.8%; Score 36; DB 2; Length 349;
 Best local Similarity 100.0%; Pred. No. 2,5e-29;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 LKGLFTLPAGHLVKEPFDITLQVNIKSTDAFPDSE 36

```

Db      152 LKGLFTLPAGHLVKEPFDTLNQVNISKTDEAFDSE 187
RESULT 4
08CCT5 PRELIMINARY; PRT; 408 AA.
AC 08CCT5;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 26, Last annotation update)
DE Mus musculus adult male olfactory brain cDNA, RIKEN full-length
   enriched library, clone:643040D06 product:hypothetical ARM repeat
   structure containing protein, full insert sequence.
DE Name:BC019693;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Mech. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
   60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
   Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Atzawa K., Nagaoka S., Sasaki N., Carninci P.,
   Kono H., Akiyama J., Nishi K., Kitanaka T., Tashiro H., Itoh M.,
   Suni N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,
   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kasaiwagi K.,
   Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
   RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RA Adachi J., Atzawa K., Akiyama T., Arakawa T., Bono H., Carninci P.,
   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
   Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
   Kato H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,

```

```

RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK032134; BAC27721.1; -
DR MGD; MGI:2384983; BC019693.
DR InterPro; IPR008938; ARM.
KM Hypothetical protein.
SQ SEQUENCE 408 AA; 4668 MW; 42AB9EB13CA3FB57 CRC64;
Query Match 87.8%; Score 36; DB 2; Length 408;
Best Local Similarity 100.0%; Pred. No. 2.8e-29;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 LKGLFTLPAGHLVKEPFDTLNQVNISKTDEAFDSE 36
Db 211 LKGLFTLPAGHLVKEPFDTLNQVNISKTDEAFDSE 246
RESULT 5
096ES5 PRELIMINARY; PRT; 349 AA.
AC 096ES5;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Flt10359 protein.
GN Name=Flt10359;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=22388257; PubMed=12477932;
RA Strauberg R.L., Feingold B.A., Grouse L.H., Derye J.G.,
   Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
   Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,
   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
   Diachenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
   Stadelton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
   Brownstein W.J., Uedin T.B., Tohiyuki S., Carninci P., Prange C.,
   Rana S.S., Loguallano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
   Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
   Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
   Whiting M., Madan A., Young A.C., Shcherenko Y., Bouffard G.G.,
   Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butcherfield Y.S.,
   Kravynets M.I., Skalska U., Smalins D.E., Schermer A., Schein J.E.,
   Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
   and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RA Strauberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC011983; AAH11983.1; -
DR InterPro; IPR008938; ARM.
SQ SEQUENCE 349 AA; 39921 MW; 3A359597FF7079EB CRC64;
Query Match 48.8%; Score 20; DB 2; Length 349;
Best Local Similarity 100.0%; Pred. No. 1.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 LKGLFTLPAGHLVKEPFDTL 20

```

DB 152 LKGLFTLFGHLYVKEPADTL 171

RESULT 6

Q8N7L7 PRELIMINARY; PRT; 897 AA.
 ID Q8N7L7
 AC Q8N7L7
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Hypothetical protein FL140893.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RX PubMed=14702039;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
 Murakami K., Yasuda T., Iwatsuki T., Waga S., Kondo H., Sugawara M.,
 Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa M., Omura Y.,
 Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
 Tanai H., Kimata M., Watanabe M., Hirooka S., Chiba Y., Ishida S.,
 Tanai H., Kimata M., Watanabe M., Hirooka S., Chiba Y., Ishida S.,
 Kanohori K., Takahashi F., Hara H., Tanase T., Kusano J.,
 Togiyama S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
 Musashino K., Yuki H., Oshima A., Sasaki N., Aotsuka S.,
 Moriya S., Motiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
 Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
 Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
 Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
 Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 Matsunuma K., Nakajima T., Mizuno T., Morinaga M., Sasaki M.,
 Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 Okumura K., Nagase T., Nomura N., Kikuchi Y., Masuho Y., Yamashita R.,
 Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
 RT "Complete sequencing and characterization of 21,243 full-length human
 cDNAs";
 RL Nat. Genet. 36:40-45 (2004).
 DR EMBL; AK098212; BAC05261.1; -
 DR InterPro; IPR008938; ARM.
 SQ SEQUENCE 897 AA; 101574 MW; 5FF6A94FE8855895 CRC64;

Query Match 48.8%; Score 20; DB 2; Length 897;

Best Local Similarity 100.0%; Pred. No. 3.2e-12;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 LKGLFTLFGHLYVKEPADTL 20
 Db 700 LKGLFTLFGHLYVKEPADTL 719

RESULT 7

Q6P197 PRELIMINARY; PRT; 1106 AA.
 ID Q6P197
 AC Q6P197
 DT 05-JUL-2004 (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE FL10359 protein (Fragment).
 GN Name=FL10359;
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 Stadelton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Udén T.B., Toohyuk S., Carrinci P., Prange C.,
 Raha S.S., Loguano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 Borak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 Richards D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 Kravinsky M.I., Skalek U., Smailus D.E., Schnerch A., Schein J.E.,
 Jones S.J., Mair M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 [2]
 SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RA Strauberg R.;
 RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC065205; AAH65205.1; -
 DR InterPro; IPR008938; ARM.
 FT NON TER 1 1
 SQ SEQUENCE 1106 AA; 125359 MW; 09F7CE940230224 CRC64;

Query Match 48.8%; Score 20; DB 2; Length 1106;
 Best Local Similarity 100.0%; Pred. No. 3.8e-12;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 LKGLFTLFGHLYVKEPADTL 20
 Db 909 LKGLFTLFGHLYVKEPADTL 928

RESULT 8

AAH65205 PRELIMINARY; PRT; 1106 AA.
 ID AAH65205
 AC AAH65205
 DT 02-MAR-2004 (TREMBlrel. 27, Created)
 DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
 DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)
 DE FL10359 protein (Fragment).
 GN FL10359;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 Stadelton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Udén T.B., Toohyuk S., Carrinci P., Prange C.,
 Raha S.A., Loguano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 Borak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 Richards D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 Kravinsky M.I., Skalek U., Smailus D.E., Schnerch A., Schein J.E.,
 Jones S.J., Mair M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 [2]
 SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RA Strauberg R.;
 RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC065205; AAH65205.1; -
 DR InterPro; IPR008938; ARM.
 FT NON TER 1 1
 SQ SEQUENCE 1106 AA; 125359 MW; 09F7CE940230224 CRC64;

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Farey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Butterfield Y.S.,
 RA Krzywinski M.I., Skalske U., Smalins D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Maiz M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RL [2]
 RN
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RA Strubeberg R.;
 DR Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC065205; AA065205.1; -.
 FT
 SQ SEQUENCE 1106 AA; 125359 MW; 09F7CE94042302C4 CRC64;
 Query Match 48.8%; Score 20; DB 2; Length 1106;
 Best Local Similarity 100.0%; Pred. No. 3,6e-12; Mismatches 0; Indels 0; Gaps 0;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 LKGLFTLFAGHLYVFPADTL 20
 Db 909 LKGLFTLFAGHLYVFPADTL 928
 RESULT 9
 Q8M9G5 PRELIMINARY; PRT; 697 AA.
 AC Q8M9G5;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE NADH dehydrogenase subunit F (Fragment).
 GN Name=ndhf;
 OS Kaliphora madagascariensis.
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Lamiales; Solanales; Montiniaceae; Kaliphora.
 OC NCBI_TaxID=125030;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bremer B., Bremer K., Heiderl N., Erixon P., Olmstead R.G.,
 RA Anderberg A.A., Kallerojo M., Barthoridian E.;
 RT "Phylogenetics of asterids based on 3 coding and 3 non-coding
 RT chloroplast DNA markers and the utility of non-coding DNA at higher
 RT taxonomic levels.";
 RT Mol. Phylogenet. Evol. 24:273-300(2002).
 RL [2]
 RN
 RP SEQUENCE FROM N.A.
 RA Lundberg J.;
 DR Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ431206; CAD34024.1; -.
 DR GO; GO:0009507; C:chloroplast; IEA.
 DR GO; GO:0009523; C:photosystem II; IEA.
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0042773; P:ATP synthesis coupled electron transport; IEA.
 DR InterPro; IPR003945; NADHb_oxrds.
 DR InterPro; IPR003916; NADHb_oxrds.
 DR InterPro; IPR001750; Oxidored_q1.
 DR InterPro; IPR002128; Oxidored_q1_C.
 DR InterPro; IPR001516; Oxidored_q1_N.
 DR Pfam; PF00361; Oxidored_q1; 1.
 DR Pfam; PF01010; Oxidored_q1_C; 1.
 DR Pfam; PF00662; Oxidored_q1_N; 1.
 DR PRINTS; PRO1434; NADHGNASES.
 DR PRINTS; PRO1435; NADHGNASES.
 DR TIGRFAMs; TIGR01974; NDH_I_L; 1.
 KM Chloroplast; NAD; NADP; Oxidoreductase; Plastoguinone; Quinone.

FT NON TER 1 1
 FT NON TER 697 697
 SQ SEQUENCE 697 AA; 78927 MW; 27E1B753AE164423 CRC64;
 Query Match 19.5%; Score 8; DB 2; Length 697;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 3 GLFTLFAG 10
 Db 547 GLFTLFAG 554
 RESULT 10
 Q9XR28 PRELIMINARY; PRT; 236 AA.
 ID Q9XR28;
 AC Q9XR28;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 05-JUN-2004 (TREMBlrel. 27, Last annotation update)
 DE MHC class II antigen.
 GN Name=IcPUDAL;
 OS Ictalurus punctatus (Channel catfish).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
 OC Ictaluridae; Ictalurus.
 OC NCBI_TaxID=7998;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20292992; PubMed=10831795;
 RA Godwin U.B., Flores M., Quinton S., Wilson M.R., Miller N.W.,
 RA Clem L.W., McConnell T.J.;
 RT "MHC class II A genes in the channel catfish (Ictalurus punctatus).";
 RL Dev. Comp. Immunol. 24:609-622(2000).
 DR EMBL; AF103003; AAD39866.1; -.
 DR HSSP; P28068; IHDH.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0006955; P:immune response; IEA.
 DR InterPro; IPR007110; Ig_I-like.
 DR InterPro; IPR003597; Ig_C1.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR001003; MHC_II_alpha.
 DR Pfam; PF07654; C1-set; 1.
 DR Pfam; PF00993; MHC_II_alpha; 1.
 DR SMART; SM00407; IGc1; 1.
 DR PROSITE; PS50835; IG_LIKE; 1.
 DR PROSITE; PS00290; IG_MHC; UNKNOWN 1.
 SQ SEQUENCE 236 AA; 26312 MW; 8C8DAB02C4CE827B CRC64;
 Query Match 17.1%; Score 7; DB 2; Length 236;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 15 PFADTLN 21
 Db 62 PFADTLN 68
 RESULT 11
 Q81DM8 PRELIMINARY; PRT; 245 AA.
 ID Q81DM8;
 AC Q81DM8;
 DT 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Zinc uptake transporter.
 GN ORFNames=BC2329;
 OS Bacillus cereus (strain ATCC 14579 / DSM 31).
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OC NCBI_TaxID=226900;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=22608415; PubMed=12721630; DOI=10.1038/nature01582;
 Ivanova N., Sorochan A., Anderson I., Galleon N., Candelson B.,
 RA Kapratel V., Bhattacharya A., Reznik G., Mikhailova N., Lapides A.,
 RA Chu L., Mazur M., Goldman E., Larsen N., D'Souza M., Malunas T.,
 RA Grechkin Y., Pusch G., HaseelKorn R., Fomstein M., Ehrlich S.D.,
 RA Overbeek R., Kyprides N.C.;
 RT "Genome sequence of *Bacillus cereus* and comparative analysis with
 RT *Bacillus anthracis*.";
 RL Nature 423:87-91(2003).
 DR EMBL; AF017005; AAP09293.1; -
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0046873; F:metal ion transporter activity; IEA.
 DR GO; GO:0030001; P:metal ion transport; IEA.
 DR InterPro; IPR003689; Zn_transp_Zip.
 DR Pfam; PF02535; Zip; 1; Zn_transp_Zip.
 SO SEQUENCE 245 AA; 27087 MW; EF5CB33CEP89C934 CRC64;

Query Match 17.1%; Score 7; DB 2; Length 245;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 FTPLFAG 11
 Db 238 FTPLFAG 244

RESULT 12
 ID 087292 PRELIMINARY; PRT; 273 AA.
 AC 087292;
 DT 01-JUN-2003 (TRENBLrel. 24, Created)
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Bacterial transferase, hexapeptide repeat protein.
 GN OrderedLocustNames=ESP703538;
 OS Pseudomonas syringae (pv. tomato).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 NCBI_TaxID=323;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DC3000;
 RA MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;
 RA Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,
 RA Gwin L., Dodson R.J., DeBoy R.T., Durkin A.S., Kolonay J.F.,
 RA Madupu R., Daugherty S.C., Brinkac L.M., Beanan M.J., Haft D.H.,
 RA Nelson W.C., Davidson T.M., Zafar N., Zhou L., Liu J., Yuan Q.,
 RA Kouri H.M., Fedorova N.B., Tran B., Russell D., Berry K.J.,
 RA Utecherback T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,
 RA Deng W.-L., Ramos A.R., Alfano J.R., Cartinhour S., Chatterjee A.K.,
 RA Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,
 RA Bender C.L., White O., Fraser C.M., Collier A.;
 RT "The complete genome sequence of the Arabidopsis and tomato pathogen
 RT *Pseudomonas syringae* pv. *tomato* DC3000.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).
 DR EMBL; AE016868; AA057013.1; -
 DR TIGR; PSP03538; -
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR InterPro; IPR001451; Hexaped_tranf.
 DR InterPro; IPR01004; Trimer_LpXA_like.
 DR Pfam; PF00132; Hexaped; 4.
 DR Complete proteome; Transferase.
 SO SEQUENCE 273 AA; 30490 MW; 6005A00571620D12 CRC64;

Query Match 17.1%; Score 7; DB 2; Length 273;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 KGLFTLF 8
 Db 266 KGLFTLF 272

RESULT 13
 ID 09A06 PRELIMINARY; PRT; 296 AA.
 AC 09A06;
 DT 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
 DE Hypothetical protein CC0498.
 GN OrderedLocustNames=CC0498;
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
 OC Caulobacteraceae; Caulobacter.
 NCBI_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SRRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647; DOI=10.1073/pnas.061029298;
 RA Niernan W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Kouri H.M., Shetty J.,
 RA Berry K.J., Utecherback T.R., Tran K., Wolf A.M., Vamathevan J.J.,
 RA Ermolaeva M.D., White O., Salzberg S.L., Venter J.C., Shapiro L.,
 RA Fraser C.M.;
 RT "Complete genome sequence of *Caulobacter crescentus*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL; AE005722; AK22485.1; -
 DR PIR; A87311; A87311.
 DR TIGR; CC0498; -
 DR InterPro; IPR003169; GYF.
 DR InterPro; IPR010432; RBD.
 DR Pfam; PF06271; RBD; 1.
 KW Complete proteome; Hypothetical protein.
 SO SEQUENCE 296 AA; 32415 MW; B9D92D2D8A921156 CRC64;

Query Match 17.1%; Score 7; DB 2; Length 296;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 LFTPLFAG 10
 Db 240 LFTPLFAG 246

RESULT 14
 ID 08HS77 PRELIMINARY; PRT; 366 AA.
 AC 08HS77;
 DT 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE NADH dehydrogenase (Fragment).
 GN Name=ndhF;
 OS Corchorus capsularis.
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eustosids II; Malvales; Malvaceae; Grewioideae; Corchorus.
 NCBI_TaxID=210143;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Whitlock B.A., Karol K.G., Alverson W.S.;
 RT "Chloroplast DNA sequences confirm the placement of the enigmatic
 RT *Oceanopappus* within *Corchorus* (Grewioideae: Malvaceae s.l., formerly
 RT *Tiliaceae*).";
 RL Int. J. Plant Sci. 164:35-41(2003).
 DR EMBL; AF523838; AA016032.1; -
 DR GO; GO:0009507; C:chloroplast; IEA.
 DR GO; GO:0009523; C:photosystem II; IEA.
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0042773; P:ATP synthesis coupled electron transport; IEA.

DR InterPro; IPR003945; NADHpl_oxrds.
 DR InterPro; IPR001750; Oxidored_q1.
 DR InterPro; IPR002128; Oxidored_q1_C.
 DR Pfam; PF00361; Oxidored_q1; 1.
 DR PRINTS; PF01010; Oxidored_q1_C; 1.
 DR PRINTS; PF01435; NPOXDRDTASE5.
 KW Chloroplast; NAD; NADP; Oxidoreductase; Plastoguinone; Quinone.
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 366 AA; 41625 MW; A6E391ABF5CA2A6A CRC64;

Query Match 17.1%; Score 7; DB 2; Length 366;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 LEFTFAG 10
 Db 229 LEFTFAG 235

RESULT 15

Q7RDUS PRELIMINARY; PRT; 367 AA.
 AC Q7RDUS;
 DT 01-MAR-2004 (TREMBLrel. 26, Created)
 DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Hypothetical protein.
 GN Name=PY05325;
 OS Plasmodium yoelii yoelii.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=73239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=17XNL;
 RX PubMed=12368865;
 RA Carlton J.M., Angluoli S.V., Suh B.B., Kooij T.W., Pertea M.,
 RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
 RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
 RA Cho J.K., Quackenbush J., Sedegah W., Shoaihi A., Cummings L.M.,
 RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Valdivia A.B.,
 RA van Lan L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
 RA Carucci D.J.;
 RT "Genome sequence and comparative analysis of the model rodent malaria
 parasite Plasmodium yoelii yoelii.";
 RL Nature 419:512-519(2002).
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AABL01001679; EAA17333.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 367 AA; 42633 MW; F6C1CD4DB0636487 CRC64;

Query Match 17.1%; Score 7; DB 2; Length 367;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 NOVNIISK 27
 Db 226 NOVNIISK 232

Search completed: February 1, 2005, 15:31:04
 Job time : 70.5556 secs

This Page Blank (uspto)

}

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:58:28 ; Search time 17.5556 Seconds
(without alignments)
154.882 Million cell updates/sec

Title: SEQ5ASP1967
Perfect score: 41
Sequence: 1 LKGFLLFAGHLVKKFPADTL.....QVNISKTDPAFFDSNDPEK 41

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size : 0

Total number of hits satisfying chosen parameters: 290433

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTCUTS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	17.1	220	4 US-09-823-823-6	Sequence 6, Appli
2	7	17.1	587	4 US-09-252-991A-31479	Sequence 31479, A
3	7	17.1	1045	1 US-07-596-467-6	Sequence 6, Appli
4	7	17.1	1045	1 US-07-934-374-6	Sequence 6, Appli
5	7	17.1	1045	1 US-07-783-861C-6	Sequence 6, Appli
6	6	14.6	41	3 US-08-630-915A-79	Sequence 79, Appli
7	6	14.6	41	4 US-09-879-957-79	Sequence 79, Appli
8	6	14.6	65	4 US-09-543-681A-7883	Sequence 7883, Ap
9	6	14.6	70	4 US-09-270-767-61018	Sequence 61018, A
10	6	14.6	120	4 US-09-710-279-220	Sequence 220, App
11	6	14.6	121	3 US-09-134-001C-3156	Sequence 3156, Ap
12	6	14.6	129	4 US-09-673-898-10	Sequence 10, Appli
13	6	14.6	160	4 US-09-328-352-6593	Sequence 6593, Ap
14	6	14.6	182	2 US-08-698-805-4	Sequence 4, Appli
15	6	14.6	200	4 US-09-222-938A-40	Sequence 40, Appli
16	6	14.6	200	4 US-09-270-767-61729	Sequence 61729, A
17	6	14.6	201	4 US-09-270-767-32397	Sequence 32397, A
18	6	14.6	201	4 US-09-270-767-47614	Sequence 47614, A
19	6	14.6	214	3 US-08-753-159A-1	Sequence 1, Appli
20	6	14.6	214	2 US-09-133-735-1	Sequence 1, Appli
21	6	14.6	223	4 US-09-583-110-3203	Sequence 3203, Ap
22	6	14.6	227	4 US-09-328-352-8175	Sequence 8175, Ap
23	6	14.6	227	4 US-09-489-039A-12794	Sequence 12794, A
24	6	14.6	230	4 US-09-252-991A-20963	Sequence 20963, A
25	6	14.6	236	4 US-09-489-039A-11254	Sequence 11254, A
26	6	14.6	240	4 US-09-134-000C-5098	Sequence 5098, Ap
27	6	14.6	242	4 US-09-252-991A-24211	Sequence 24211, A

28	6	14.6	244	4 US-09-252-991A-25108	Sequence 25108, A
29	6	14.6	251	4 US-09-489-039A-8708	Sequence 8708, Ap
30	6	14.6	272	4 US-09-489-039A-7644	Sequence 7644, Ap
31	6	14.6	313	4 US-09-248-786A-18880	Sequence 18880, A
32	6	14.6	323	4 US-09-538-032-213	Sequence 213, App
33	6	14.6	329	4 US-09-651-200-19	Sequence 19, Appli
34	6	14.6	331	4 US-09-830-217-18	Sequence 851, Appli
35	6	14.6	349	4 US-09-198-452A-851	Sequence 851, App
36	6	14.6	361	3 US-09-261-599B-4	Sequence 4, Appli
37	6	14.6	361	4 US-09-456-455A-4	Sequence 4, Appli
38	6	14.6	388	4 US-09-248-786A-18781	Sequence 18781, A
39	6	14.6	414	4 US-09-388-050-6	Sequence 6, Appli
40	6	14.6	417	4 US-09-270-767-46168	Sequence 46168, A
41	6	14.6	447	4 US-09-540-236-3675	Sequence 3675, Ap
42	6	14.6	447	4 US-09-388-089B-2	Sequence 2, Appli
43	6	14.6	455	4 US-09-248-786A-17535	Sequence 17535, A
44	6	14.6	465	4 US-09-388-050-4	Sequence 4, Appli
45	6	14.6	473	4 US-09-252-991A-29884	Sequence 29884, A

ALIGNMENTS

```

RESULT 1
US-09-823-823-6
; Sequence 6, Application US/09823823
; Patent No. 6635904
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Satoshi
; APPLICANT: Kasai, Hiroaki
; APPLICANT: Nakamura, Shoko
; APPLICANT: Suzuki, Makoto
; APPLICANT: Hamada, Toshiyuki
; TITLE OF INVENTION: METHOD FOR IDENTIFICATION AND DETECTION OF MICROORGANISMS USING G
; FILE REFERENCE: 12817-004001
; CURRENT FILING DATE: US/09/823, 823
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US 09/208, 688
; PRIOR FILING DATE: 1998-12-10
; PRIOR APPLICATION NUMBER: JP 97/343316
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 2.0
; SEQ ID NO 6
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Chitinophaga pinensis
US-09-823-823-6
;
;
Query Match      17.1%; Score 7; DB 4; Length 220;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches          7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY              35 SENDPEK 41
DB              130 SENDPEK 136

RESULT 2
US-09-252-991A-31479
; Sequence 31479, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT FILING DATE: US/09/252, 991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074, 788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094, 190

```

PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 31479
LENGTH: 587
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31479

Query Match 17.1%; Score 7; DB 4; Length 587;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ADTLDGV 23
DB 381 ADTLDGV 387

RESULT 3
US-07-596-467-6
Sequence 6, Application US/07596467
Patent No. 5306862
GENERAL INFORMATION:

APPLICANT: Chappell, J.
APPLICANT: Saunders, Court A.
APPLICANT: Cuellar, R.
TITLE OF INVENTION: Method and Composition for Increasing
TITLE OF INVENTION: Sterol Accumulation in Higher Plants
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:

ADDRESSER: Dressler, Goldsmith, Shore, Sutker & Milnamow
STREET: 180 N. Stetson St.
CITY: Chicago
STATE: Illinois
COUNTRY: USA

ZIP: 60601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/596,467

FILING DATE: 19901012

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Gamson, Edward P.

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312 616-5400

TELEFAX: 312 616-5460

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 1045 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-596-467-6

QY 11 HLKPPA 17

DB 10 HLKPPA 16

RESULT 4
US-07-934-374-6
Sequence 6, Application US/07934374
Patent No. 5349126
GENERAL INFORMATION:
APPLICANT: Chappell, J.

APPLICANT: Saunders, Court A.
APPLICANT: Cuellar, R.
APPLICANT: Wolf, Fred R.
TITLE OF INVENTION: Process and Composition for Increasing
TITLE OF INVENTION: Sterol Accumulation in Higher Plants
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:

ADDRESSER: Dressler, Goldsmith, Shore, Sutker & Milnamow
STREET: 180 N. Stetson St.
CITY: Chicago
STATE: Illinois
COUNTRY: USA

ZIP: 60601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/934,374

FILING DATE: 19920814

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Gamson, Edward P.

REGISTRATION NUMBER: 29,381

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312 616-5400

TELEFAX: 312 616-5460

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 1045 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-934-374-6

QY 11 HLKPPA 17

DB 10 HLKPPA 16

RESULT 5
US-07-783-861C-6

Sequence 6, Application US/07783861C

Patent No. 5460949

GENERAL INFORMATION:

APPLICANT: Saunders, Court A.

APPLICANT: Wolf, Fred R.

APPLICANT: Mukharji, Indrani

TITLE OF INVENTION: A Method and Composition for Increasing

TITLE OF INVENTION: the Accumulation of Squalene and Specific Sterols in

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSER: Amoco Corp., Patents and Licensing Dept.

STREET: 200 East Randolph St.

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60680-0703

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/783,861C

FILING DATE: 19911028

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/613,380
FILING DATE: 15-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Galloway, No. 5460949v11 B.
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312 856-7180
TELEFAX: 312 856-4972
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1045 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-783-861C-6

Query Match 17.1%; Score 7; DB 1; Length 1045;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 11 HLVKPA 17
Db 10 HLVKPA 16

RESULT 6

US-08-630-915A-79
Sequence 79, Application US/08630915A

PATENT No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, No. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOWLES, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESS: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 41 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown

MOLECULE TYPE: peptide
US-08-630-915A-79

Query Match 14.6%; Score 6; DB 3; Length 41;

Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 36 ENDPEK 41
Db 29 ENDPEK 34

RESULT 7

US-09-879-957-79
Sequence 79, Application US/09879957

PATENT No. 6709821
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, No. 6709821h
APPLICANT: KAY, Brian K.
APPLICANT: FOWLES, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESS: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 41 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: unknown

MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-09-879-957-79

Query Match 14.6%; Score 6; DB 4; Length 41;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 36 ENDPEK 41
Db 29 ENDPEK 34

RESULT 8

US-09-543-681A-7883
Sequence 7883, Application US/09543681A

PATENT No. 6605709
GENERAL INFORMATION:

```

; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7883
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7883

Query Match
Best Local Similarity 14.6%; Score 6; DB 4; Length 65;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 VNISK 28
DB 55 VNISK 60

RESULT 9
US-09-270-767-61018
; Sequence 61018, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 61018
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-61018

Query Match
Best Local Similarity 14.6%; Score 6; DB 4; Length 70;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 BAFPS 35
DB 47 BAFPS 52

RESULT 10
US-09-710-279-220
; Sequence 220, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 220
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic

```

```

; OTHER INFORMATION: amino acid sequence
US-09-710-279-220

Query Match
Best Local Similarity 14.6%; Score 6; DB 4; Length 120;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 VNISK 28
DB 81 VNISK 86

RESULT 11
US-09-134-001C-3156
; Sequence 3156, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3156
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3156

Query Match
Best Local Similarity 14.6%; Score 6; DB 3; Length 121;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 VNISK 28
DB 82 VNISK 87

RESULT 12
US-09-673-898-10
; Sequence 10, Application US/09673898
; Patent No. 6709657
; GENERAL INFORMATION:
; APPLICANT: Ruelle, Jean Louis
; TITLE OF INVENTION: BASHO13 DNA AND PROTEINS FROM NEISSERIA
; FILE REFERENCE: BM45307
; CURRENT APPLICATION NUMBER: US/09/673,898
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/EP99/02765
; PRIOR FILING DATE: 1999-04-20
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 129
; TYPE: PRT
; ORGANISM: Bacteria
US-09-673-898-10

Query Match
Best Local Similarity 14.6%; Score 6; DB 4; Length 129;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 SKTDEA 31
DB 31 SKTDEA 36

```

```
RESULT 13
US-09-328-352-6593
; Sequence 6593, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6593
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6593

Query Match
Best Local Similarity 14.6%; Score 6; DB 4; Length 160;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      12 LVKPFA 17
Db      121 LVKPFA 126

RESULT 14
US-08-698-805-4
; Sequence 4, Application US/08698805
; Patent No. 5869288
; GENERAL INFORMATION:
; APPLICANT: Chapman, Martin
; APPLICANT: Arruda, L. Karla
; TITLE OF INVENTION: Molecular Cloning of Cockroach
; TITLE OF INVENTION: Allele, Amino Acid and Nucleotide Sequences Therefore,
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oblon, Spivak, McCIelland, Moler & Neustadt,
; ADDRESSSEE: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent'n Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/698,805
; FILING DATE: 16-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/002,510
; FILING DATE: 18-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Kelder, Steven B.
; REGISTRATION NUMBER: 30,073
; REFERENCE/DOCKET NUMBER: 494-203-27
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-5000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 182 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-698-805-4
```

```
Query Match
Best Local Similarity 14.6%; Score 6; DB 2; Length 182;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      19 TLDOVN 24
Db      152 TLDOVN 157

RESULT 15
US-09-222-938A-40
; Sequence 40, Application US/09222938A
; Patent No. 6437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Filtz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 0734/060001
; CURRENT APPLICATION NUMBER: US/09/222,938A
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 40
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-222-938A-40

Query Match
Best Local Similarity 14.6%; Score 6; DB 4; Length 200;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      22 QVNISK 27
Db      121 QVNISK 126

Search completed: February 1, 2005, 15:35:44
Job time : 18.5556 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 15:09:39 ; Search time 59.7778 Seconds
(without alignment)
247.799 Million cell updates/sec

Title: SEQ5ASPI967

Perfect score: 41
Sequence: 1 LKGLFTLFAHLVLPFADTL.....QVNISKTEAFDSENDPEK 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1608061 seqs, 361289386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1311380

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
- 12: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubppaa/US10D_PUBCOMB.pep:*
- 17: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubppaa/US11_NEW_PUB.pep:*
- 19: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
- 20: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	48.8	1149	17 US-10-128-558-167	Sequence 167, App
2	7	17.1	182	16 US-10-437-963-193489	Sequence 193489
3	7	17.1	220	9 US-09-823-829-6	Sequence 6, Appl
4	7	17.1	220	9 US-09-823-823-6	Sequence 6, Appl
5	7	17.1	1045	14 US-10-369-493-1836	Sequence 1836, App
6	7	17.1	1045	15 US-10-041-018-199	Sequence 199, App
7	7	17.1	1045	15 US-10-041-018-218	Sequence 218, App
8	7	17.1	1045	15 US-10-041-018-264	Sequence 264, App
9	7	17.1	1045	15 US-10-041-018-309	Sequence 309, App
10	7	17.1	1045	15 US-10-041-018-336	Sequence 336, App
11	6	14.6	41	9 US-09-879-957-79	Sequence 79, Appl
12	6	14.6	41	16 US-10-807-856-79	Sequence 79, Appl
13	6	14.6	45	9 US-09-864-761-34609	Sequence 34609, A

14	6	14.6	51	9 US-09-864-761-47882	Sequence 47882, A
15	6	14.6	55	15 US-10-424-599-284582	Sequence 284582
16	6	14.6	58	16 US-10-437-963-104980	Sequence 104980
17	6	14.6	69	16 US-10-767-701-53471	Sequence 53471, A
18	6	14.6	78	17 US-10-425-115-277649	Sequence 277649
19	6	14.6	80	15 US-10-424-599-143186	Sequence 143186
20	6	14.6	80	17 US-10-425-115-237591	Sequence 237591
21	6	14.6	87	17 US-10-425-115-250582	Sequence 250582
22	6	14.6	91	15 US-10-424-599-269946	Sequence 269946
23	6	14.6	93	15 US-10-424-599-183753	Sequence 183753
24	6	14.6	95	16 US-10-437-963-114188	Sequence 114188
25	6	14.6	95	16 US-10-767-701-46828	Sequence 46828, A
26	6	14.6	99	15 US-10-424-599-205004	Sequence 205004
27	6	14.6	105	15 US-10-424-599-157330	Sequence 157330
28	6	14.6	111	15 US-10-424-599-244159	Sequence 244159
29	6	14.6	111	17 US-10-425-115-340045	Sequence 340045
30	6	14.6	115	15 US-10-041-018-211	Sequence 211, App
31	6	14.6	115	15 US-10-041-018-351	Sequence 351, App
32	6	14.6	118	15 US-10-424-599-177366	Sequence 177366
33	6	14.6	125	16 US-10-437-963-114015	Sequence 114015
34	6	14.6	127	16 US-10-767-701-56103	Sequence 56103, A
35	6	14.6	129	16 US-10-739-956-10	Sequence 10, Appl
36	6	14.6	130	17 US-10-425-115-198836	Sequence 198836
37	6	14.6	133	17 US-10-425-115-194350	Sequence 194350
38	6	14.6	143	17 US-10-425-115-342272	Sequence 342272
39	6	14.6	144	9 US-09-764-878-127	Sequence 127, App
40	6	14.6	144	9 US-09-764-860-461	Sequence 461, App
41	6	14.6	144	14 US-10-079-854-127	Sequence 127, App
42	6	14.6	144	14 US-10-074-095-461	Sequence 461, App
43	6	14.6	144	14 US-10-212-872-461	Sequence 461, App
44	6	14.6	150	17 US-10-425-115-351501	Sequence 351501
45	6	14.6	171	15 US-10-424-599-148350	Sequence 148350

ALIGNMENTS

RESULT 1
US-10-128-558-167
Sequence 167, Application US/10128558
Publication No. US20040219521A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Wang, Zhiwei
APPLICANT: Wang, Gezhi
APPLICANT: Boyle, Bryan J
APPLICANT: Drmanac, Radjoe T
TITLE OF INVENTION: Novel Nucleic Acids and
FILE REFERENCE: 812A
CURRENT APPLICATION NUMBER: US/10/128,558
CURRENT FILING DATE: 2002-04-22
PRIOR APPLICATION NUMBER: US 60/339,453
PRIOR FILING DATE: 2001-12-11
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02823
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: PCT/US01/03800
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 09/515,126
PRIOR FILING DATE: 2000-02-28
Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 412
SOFTWARE: pc_Fl_genes Version 6.0
SEQ ID NO 167
LENGTH: 1149
TYPE: PRT
ORGANISM: Homo sapiens
US-10-128-558-167

Query Match 48.8%; Score 20; DB 17; Length 1149;
Best Local Similarity 100.0%; Pred. No. 6,8e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LKGLFTLFGHLYKPPADTL 20
DB 952 LKGLFTLFGHLYKPPADTL 971

RESULT 2

US-10-437-963-193489
Sequence 193489, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Bardazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants and Uses thereof for Plant Improvement
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 193489
LENGTH: 182
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(182)
OTHER INFORMATION: unsure at all Xaa locations
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MKT4530_89622C.1.pcp
US-10-437-963-193489

Query Match 17.1%; Score 7; DB 16; Length 182;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LKGLFTL 7
DB 19 LKGLFTL 25

RESULT 3

US-09-823-829-6
Sequence 6, Application US/09823829
Patent No. US20020146697A1
GENERAL INFORMATION:
APPLICANT: Yamamoto, Satoshi
APPLICANT: Nakamura, Shoko
APPLICANT: Suzuki, Makoto
APPLICANT: Kasai, Hiroaki
APPLICANT: Hamada, Tohtu
TITLE OF INVENTION: METHOD FOR IDENTIFICATION AND DETECTION OF MICROORGANISMS
TITLE OF INVENTION: USING GYRASE GENE AS AN INDICATOR
FILE REFERENCE: 12817-004001
CURRENT APPLICATION NUMBER: US/09/823,829
CURRENT FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: US 09/208,668

PRIOR FILING DATE: 1998-12-10
PRIOR APPLICATION NUMBER: JP 97/343316
PRIOR FILING DATE: 1997-12-12
NUMBER OF SEQ ID NOS: 82
SOFTWARE: PatentIn version 2.0
SEQ ID NO 6
LENGTH: 220
TYPE: PRT
ORGANISM: Chitinophaga pinensis
US-09-823-829-6

Query Match 17.1%; Score 7; DB 9; Length 220;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 SENDPEK 41
DB 130 SENDPEK 136

RESULT 4

US-09-823-823-6
Sequence 6, Application US/09823823
Patent No. US20020171092A1
GENERAL INFORMATION:
APPLICANT: Yamamoto, Satoshi
APPLICANT: Kasai, Hiroaki
APPLICANT: Nakamura, Shoko
APPLICANT: Suzuki, Makoto
APPLICANT: Hamada, Tohtu
TITLE OF INVENTION: METHOD FOR IDENTIFICATION AND DETECTION OF MICROORGANISMS USING GY
TITLE OF INVENTION: GENE AS AN INDICATOR
FILE REFERENCE: 12817-004001
CURRENT APPLICATION NUMBER: US/09/823,823
CURRENT FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: US 09/208,688
PRIOR FILING DATE: 1998-12-10
PRIOR APPLICATION NUMBER: JP 97/343316
PRIOR FILING DATE: 1997-12-12
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn version 2.0
SEQ ID NO 6
LENGTH: 220
TYPE: PRT
ORGANISM: Chitinophaga pinensis
US-09-823-823-6

Query Match 17.1%; Score 7; DB 9; Length 220;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 SENDPEK 41
DB 130 SENDPEK 136

RESULT 5

US-10-369-493-1836
Sequence 1836, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianteng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21

```
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1836
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-1836
```

```
Query Match          17.1%; Score 7; DB 14; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      11 HLVPKFA 17
        |||||
Db      10 HLVPKFA 16
```

```
RESULT 6
US-10-041-018-199
; Sequence 199, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041.018
; PRIOR FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 199
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-041-018-199
```

```
Query Match          17.1%; Score 7; DB 15; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      11 HLVPKFA 17
        |||||
Db      10 HLVPKFA 16
```

```
RESULT 7
US-10-041-018-218
; Sequence 218, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041.018
; PRIOR FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 218
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-041-018-218
```

```
Query Match          17.1%; Score 7; DB 15; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      11 HLVPKFA 17
        |||||
```

```
Db      10 HLVPKFA 16
```

```
RESULT 8
US-10-041-018-264
; Sequence 264, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041.018
; PRIOR FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 264
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-041-018-264
```

```
Query Match          17.1%; Score 7; DB 15; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      11 HLVPKFA 17
        |||||
Db      10 HLVPKFA 16
```

```
RESULT 9
US-10-041-018-309
; Sequence 309, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041.018
; PRIOR FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 309
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-041-018-309
```

```
Query Match          17.1%; Score 7; DB 15; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      11 HLVPKFA 17
        |||||
Db      10 HLVPKFA 16
```

```
RESULT 10
US-10-041-018-336
; Sequence 336, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041.018
```

CURRENT FILING DATE: 2002-01-07
PRIOR APPLICATION NUMBER: US 60/259880
PRIOR FILING DATE: 2001-01-05
NUMBER OF SEQ ID NOS: 413
SOFTWARE: Patentin version 3.1
SEQ ID NO 336
LENGTH: 1045
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-10-041-018-336

Query Match 17.1%; Score 7; DB 15; Length 1045;
Best Local Similarity 100.0%; Pred.No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 HLVKPPA 17
Db 10 HLVKPPA 16

RESULT 11

US-09-879-957-79
Sequence 79, Application US/09879957
Patent No. US20020034755A1
GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.
HOFFMAN, No. US20020034755A1h
KAY, Brian K.
FOWLES, Dana M.
MCCONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/879,957

FILING DATE: 13-Jun-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/630,915

FILING DATE: 03-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-174

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864/9741

INFORMATION FOR SEQ ID NO: 79:

SEQUENCE CHARACTERISTICS:

LENGTH: 41 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: unknown

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 79:

Query Match 14.6%; Score 6; DB 9; Length 41;
Best Local Similarity 100.0%; Pred.No. 89;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 ENDPEK 41
Db 29 ENDPEK 34

RESULT 12

US-10-807-856-79
Sequence 79, Application US/10807856
Publication No. US20040157216A1
GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.
HOFFMAN, Noah
KAY, Brian K.
FOWLES, Dana M.
MCCONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/807,856

FILING DATE: 23-Mar-2004

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/630,915

FILING DATE: 03-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-174

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864/9741

INFORMATION FOR SEQ ID NO: 79:

SEQUENCE CHARACTERISTICS:

LENGTH: 41 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: unknown

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 79:

Query Match 14.6%; Score 6; DB 16; Length 41;
Best Local Similarity 100.0%; Pred.No. 89;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 ENDPEK 41
Db 29 ENDPEK 34

RESULT 13

US-09-864-761-34609
Sequence 34609, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1
CURRENT FILING DATE: 2001-05-23
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-02-04
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 34609
LENGTH: 45
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC007383.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.8
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.6
OTHER INFORMATION: EXPRESSED IN HEAT, SIGNAL = 9.8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1
OTHER INFORMATION: SWISSPROT HIT: P28331, EVALUATE 9.00e-20
OTHER INFORMATION: EST_HUMAN HIT: AW614813.1, EVALUATE 3.00e-19
US-09-864-761-34609

Query Match 14.6%; Score 6; DB 9; Length 45;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 18 DTLDOV 23
DB 9 DTLDOV 14

RESULT 14
US-09-864-761-47882
Sequence 47882, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1
CURRENT FILING DATE: 2001-05-23
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 47882
LENGTH: 51
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC008897.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.1
OTHER INFORMATION: SWISSPROT HIT: A0132001.1, EVALUATE 7.00e-23
OTHER INFORMATION: EST_HUMAN HIT: P04035, EVALUATE 6.00e-24
US-09-864-761-47882

Query Match 14.6%; Score 6; DB 9; Length 51;
Best Local Similarity 100.0%; Pred. No. 11e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 AGHLVK 14
DB 40 AGHLVK 45

RESULT 15

US-10-424-599-284582
; Sequence 284582, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 284582
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_99002C.1.pap
US-10-424-599-284582

Query Match 14.6%; Score 6; DB 15; Length 55;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 HUVKPF 16
Db 30 HUVKPF 35

Search completed: February 1, 2005, 15:44:50
Job time : 61.7778 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:55:01 ; Search time 13.5556 Seconds
(without alignments)
291.016 Million cell updates/sec

Title: SEQ5ASP1967

Perfect score: 41
Sequence: 1 LKGFPLFAGHLVLPFADTL.....QVNISKTEAEPFSENDPEK 41Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 279530

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	17.1	296	2 A87311	hypothetical prote
2	7	17.1	400	2 C82213	conserved hypotnet
3	7	17.1	422	2 D87550	hypothetical prote
4	7	17.1	453	2 F86846	ABC transporter pe
5	7	17.1	1045	2 B30239	hydroxymethylgluta
6	6	14.6	75	2 F75451	hypothetical prote
7	6	14.6	107	2 E82494	conserved hypotnet
8	6	14.6	120	2 E89932	conserved hypotnet
9	6	14.6	131	2 C83863	hypothetical prote
10	6	14.6	134	2 G69947	phage-related prot
11	6	14.6	146	2 H72527	hypothetical prote
12	6	14.6	162	2 T17044	NADH2 dehydrogenas
13	6	14.6	173	2 S59507	ferric pseudobacti
14	6	14.6	183	2 B75152	adenylyl cyclase r
15	6	14.6	183	2 C71193	hypothetical prote
16	6	14.6	201	2 B84671	hypothetical prote
17	6	14.6	210	2 H84161	hypothetical prote
18	6	14.6	212	2 A83065	conserved hypotnet
19	6	14.6	212	2 F88221	ycyf protein limpo
20	6	14.6	219	2 T38019	uracil phosphoribo
21	6	14.6	220	2 A75287	response regulator
22	6	14.6	220	2 A00426	two-component byst
23	6	14.6	221	2 T16578	hypothetical prote
24	6	14.6	222	1 B40656	regulatory protein
25	6	14.6	222	1 H65220	transcription regu
26	6	14.6	222	2 AD1022	two-component resp
27	6	14.6	222	2 D86106	hypothetical prote
28	6	14.6	222	2 G91265	hypothetical prote
29	6	14.6	223	2 G95193	ylme protein limpo

30	6	14.6	223	2 D98060	conserved hypotnet
31	6	14.6	225	2 C97435	transcription acti
32	6	14.6	225	2 AG2653	two component resp
33	6	14.6	226	2 G83293	probable two-compo
34	6	14.6	226	2 T35202	probable two-compo
35	6	14.6	227	1 B47080	transcription acti
36	6	14.6	227	2 F83097	probable two-compo
37	6	14.6	227	2 A90705	probable 2-compone
38	6	14.6	227	2 D85555	probable transcrip
39	6	14.6	227	2 A64780	hypothetical prote
40	6	14.6	231	2 T32953	hypothetical prote
41	6	14.6	233	2 C70128	hypothetical prote
42	6	14.6	244	2 T15807	hypothetical prote
43	6	14.6	263	1 H70046	probable 3-oxoacyl
44	6	14.6	270	2 AC1781	hypothetical prote
45	6	14.6	277	2 F82874	conserved hypotnet

ALIGNMENTS

RESULT 1

A87311 hypothetical protein CC0498 [imported] - Caulobacter crescentus

C/Species: Caulobacter crescentus

C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004

C/Accession: A87311

R/Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A/Title: Complete Genome Sequence of Caulobacter crescentus.

A/Reference number: A87249; MUID:21173698; PMID:11259647

A/Accession: A87311

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-296 <STO>

A/Cross-references: UNIPROT:Q9NAU6; GB:AE005673; NID:g13421677; PIDN:AAK2485.1; GSPDB:G

C/Genetics:

A/Gene: CC0498

Query Match 17.1%; Score 7; DB 2; Length 296;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 LFTLPAG 10
Db 240 LFTLPAG 246

RESULT 2

C82213 conserved hypothetical protein VC1339 [imported] - Vibrio cholerae (strain N16961 serogr

C/Species: Vibrio cholerae

C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004

C/Accession: C82213

R/Heidelberg, U.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.

charlson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoti, I.; Sellers, P

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A/Reference number: A82035; MUID:20406833; PMID:10952301

A/Accession: C82213

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-400 <HBI>

A/Cross-references: UNIPROT:Q9KSB9; GB:AE004213; GB:AE003852; NID:g9655819; PIDN:AAF9449

A/Experimental source: serogroup O1, strain N16961; biotype El Tor

C/Genetics:

A/Gene: VC1339

A/Map position: 1

C/Superfamily: uncharacterized conserved protein

Query Match 17.1%; Score 7; DB 2; Length 400;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 QVNIKT 28
DB 139 QVNIKT 145

RESULT 3

hypothetical protein CC2429 [imported] - Caulobacter crescentus
C/Species: Caulobacter crescentus

C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004

C/Accession: D87550

R/Nierman, W.C.; Feildbylun, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Klotz

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A/Title: Complete Genome Sequence of *Caulobacter crescentus*.

A/Reference number: A87249; MUID:21173698; PMID:11259647

A/Accession: D87550

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-422 <STO>

A/Cross-references: UNIPROT:Q9A5L8; GB:AE005673; NID:g13423972; PIDN:AAK24400.1; GSPDB:C

C/Genetics:

A/Gene: CC2429

Query Match 17.1%; Score 7; DB 2; Length 422;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LFTLPAG 10
DB 204 LFTLPAG 210

RESULT 4

ABC transporter permease protein ysdA [imported] - *Lactococcus lactis* subsp. *lactis* (str

C/Species: *Lactococcus lactis* subsp. *lactis*

C/Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004

C/Accession: F86846

R/Bolotin, A.; Winker, P.; Manger, S.; Jallion, O.; Malarne, K.; Weissenbach, J.; Ehrlich

Genome Res. 11, 731-753, 2001

A/Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ss

A/Reference number: A86625; MUID:21235186; PMID:11337471

A/Accession: F86846

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-453 <STO>

A/Cross-references: UNIPROT:Q9CER2; GB:AE005176; PID:g12724797; PIDN:AAK05872.1; GSPDB:C

C/Genetics:

A/Gene: ysdA

Query Match 17.1%; Score 7; DB 2; Length 453;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 QVNIKT 28
DB 127 QVNIKT 133

RESULT 5

hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) II - yeast (*Saccharomyces cerevisiae*)

N/Alternate names: protein U9324.2; Protein YLR450w

C/Species: *Saccharomyces cerevisiae*

C/Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 09-Jul-2004

C/Accession: B30239; B24317; S55972

R/Basson, M.E.; Thorness, M.; Finer-Moore, J.; Stroud, R.M.; Rine, J.
Mol. Cell. Biol. 8, 3797-3808, 1988

A/Title: Structural and functional conservation between yeast and human 3-hydroxy-3-methyl

A/Reference number: A93105; MUID:89127221; PMID:3065625

A/Accession: B30239

A/Molecule type: DNA

A/Residues: 1-1045 <BAS>

A/Cross-references: UNIPROT:P12684; EMBL:M22255; NID:g171687; PIDN:AAA34677.1; PID:g17168

R/Basson, M.E.; Thorness, M.; Rine, J.

Proc. Natl. Acad. Sci. U.S.A. 83, 5563-5567, 1986

A/Title: *Saccharomyces cerevisiae* contains two functional genes encoding 3-hydroxy-3-methyl

A/Reference number: A94112; MUID:96287298; PMID:3526336

A/Accession: B24317

A/Molecule type: DNA

A/Residues: 772-961 <BAS2>

Ridu, Z.

submitted to the EMBL Data Library, March 1995

A/Description: The sequence of *S. cerevisiae* cosmid 9324.

A/Reference number: S55966

A/Accession: S55972

A/Molecule type: DNA

A/Residues: 1-1045 <DU2>

A/Cross-references: GB:U2382; NID:9717059; PIDN:AA67527.1; PID:g1717061; MIPS:YLR450w

C/Genetics:

A/Gene: SCD:HM02

A/Map position: 12R

C/Superfamily: hydroxymethylglutaryl-CoA reductase I

C/Keywords: coenzyme A; oxidoreductase; sterol biosynthesis; transmembrane protein

F/29-45/Domain: transmembrane #status predicted <TM1>

F/248-264/Domain: transmembrane #status predicted <TM2>

F/331-347/Domain: transmembrane #status predicted <TM3>

F/402-418/Domain: transmembrane #status predicted <TM4>

F/502-518/Domain: transmembrane #status predicted <TM5>

F/681-697/Domain: transmembrane #status predicted <TM6>

F/991-1007/Domain: transmembrane #status predicted <TM7>

Query Match 17.1%; Score 7; DB 2; Length 1045;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 HLVKPPA 17
DB 10 HLVKPPA 16

RESULT 6

hypothetical protein - *Deinococcus radiodurans* (strain R1)

C/Species: *Deinococcus radiodurans*

C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C/Accession: F75451

R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A/Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.

A/Reference number: A75250; MUID:20036896; PMID:10567266

A/Accession: F75451

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-75 <WHI>

A/Cross-references: UNIPROT:Q9RVN1; GB:AE001951; GB:AE000513; NID:g6458712; PIDN:AAFI057;

C/Genetics:

A/Gene: DR0996

A/Map position: 1

Query Match 14.6%; Score 6; DB 2; Length 75;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LFRGHL 12

Db 20 LFAGHTL 25

RESULT 7

conserved hypothetical protein VCA0152 [imported] - Vibrio cholerae (strain N16961 serog C)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: B82494
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.; Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F. I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: B82494
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-107 <HEI>
A:Cross-references: UNIPROT:Q9KN14; GB:AE004356; GB:AE003853; NID:99657536; PIDN:AAF9606
C:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA0152
A:Map position: 2
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1428

Query Match 14.6%; Score 6; DB 2; Length 107;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 LFTLPA 9
Db 14 LFTLPA 19

RESULT 8

conserved hypothetical protein SA1356 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: B89932
R:Kuroda, M.; Ohta, Y.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramoto, K.
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: B89932
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-120 <KUN>
A:Cross-references: UNIPROT:Q99TW8; GB:BA000018; PID:g13701324; PIDN:BAB42618.1; GSPDB:C
C:Experimental source: strain N315
C:Genetics:
A:Gene: SA1356
C:Superfamily: Bacillus subtilis conserved hypothetical protein yghY

Query Match 14.6%; Score 6; DB 2; Length 120;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 23 VNISKY 28
Db 81 VNISKY 86

RESULT 9

C83863
hypothetical protein BH1707 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C:Accession: C83863

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hirata
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: C83863
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-131 <STD>
A:Cross-references: UNIPROT:Q9KC67; GB:AP001512; GB:BA000004; NID:g10174030; PIDN:BA0544
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH1707

Query Match 14.6%; Score 6; DB 2; Length 131;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 13 VKPFPAD 18
Db 45 VKPFPAD 50

RESULT 10

phage-related protein homolog yqbl - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C:Accession: G69947
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Berr
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: G69947
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-134 <KUN>
A:Cross-references: UNIPROT:P45928; GB:Z39117; GB:AL009126; NID:g2634966; PIDN:CAB14548.1
A:Experimental source: strain 168
C:Genetics:
A:Gene: yqbl

Query Match 14.6%; Score 6; DB 2; Length 134;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 33 FDSEND 38
Db 95 FDSEND 100

RESULT 11

H72527
hypothetical protein APE2197 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C:Accession: H72527
R:Kawabuchi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyru
A:Reference number: A72450; MUID:99310339; PMID:10382966

A/Accession: H72527
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-146 <KAM>
 A/Cross-references: UNIPROT:Q9Y9U2; DDBJ:AP000063; NID:G5105654; PIDN:BAA81208.1; PID:G5105654
 A/Experimental source: strain K1
 C/Genetics:
 A/Gene: AP2197
 C/Superfamily: Aeropyrum pernix hypothetical protein APE2197

Query Match 14.6%; Score 6; DB 2; Length 146;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 18 DTLDQV 23
 |||||
 Db 69 DTLDQV 74

RESULT 12

T17044
 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - common snowdrop chloroplast (fr
 C/Species: chloroplast Galanthus nivalis (Common snowdrop)
 C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C/Accession: T17044
 R/Graham, S.W.; Barrett, S.C.H.
 submitted to the EMBL Data Library, November 1996
 A/Description: Phylogenetic analysis of Narcissus L. (Amaryllidaceae) based on the chlor
 A/Reference number: Z17939
 A/Accession: T17044
 A/Status: preliminary; translated from GB/EMBL/DDBJ
 A/Molecule type: mRNA
 A/Residues: 1-162 <GRA>
 A/Cross-references: UNIPROT:O47204; EMBL:U79222; NID:G2827789; PID:G2827790; PIDN:AAB998
 C/Genetics:
 A/Genome: chloroplast
 A/Note: ndhF
 C/Superfamily: NADH dehydrogenase (ubiquinone) chain 5
 C/Keywords: chloroplast; membrane-associated complex; NAD; oxidoreductase

Query Match 14.6%; Score 6; DB 2; Length 162;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 GLFTLF 8
 |||||
 Db 63 GLFTLF 68

RESULT 13

S59507
 ferric pseudobactins receptor protein RFS - Pseudomonas putida (fragment)
 C/Species: Pseudomonas putida
 C/Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 13-Mar-1997
 C/Accession: S59507
 R/Koster, M.; Ova, W.; Bitter, W.; Weisbeek, P.
 Mol. Gen. Genet. 248, 735-743, 1995
 A/Title: Multiple outer membrane receptors for uptake of ferric pseudobactins in Pseudom
 A/Reference number: S59503; MUID:96069713; PMID:7476877
 A/Accession: S59507
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-173 <KOS>

Query Match 14.6%; Score 6; DB 2; Length 173;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 GLFTLF 8
 |||||
 Db 98 GLFTLF 103

RESULT 14

B75152
 adenyl cyclase related protein PAB2098 - Pyrococcus abyssi (strain Orsay)
 C/Species: Pyrococcus abyssi
 C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 16-Aug-2004
 C/Accession: B75152
 R/anonymous, Genoscope
 submitted to the EMBL Data Library, July 1999
 A/Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome struc
 A/Reference number: A75001
 A/Accession: B75152
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-183 <KAM>
 A/Cross-references: UNIPROT:Q9V102; GB:AJ248284; GB:AL096836; NID:G5457730; PIDN:CAB4929;
 A/Experimental source: strain Orsay
 C/Genetics:
 A/Gene: PAB2098
 C/Superfamily: Thermophilic adenylate cyclase, Cyab type

Query Match 14.6%; Score 6; DB 2; Length 183;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 26 SKTDEA 31
 |||||
 Db 43 SKTDEA 48

RESULT 15

C71193
 hypothetical protein PH1819 - Pyrococcus horikoshii
 C/Species: Pyrococcus horikoshii
 C/Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
 C/Accession: C71193
 R/Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hata, Y.; Hino, Y.; Yamamoto, S.; Sekine
 M.; Ohnuki, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
 DNA Res. 5, 55-76, 1998
 A/Title: Complete sequence and gene organization of the genome of a hyper-thermophilic ar
 A/Reference number: A71000; MUID:98344137; PMID:9679194
 A/Accession: C71193
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-183 <KAM>
 A/Cross-references: UNIPROT:O59483; GB:AP000007; NID:G3236134; PIDN:BAA30938.1; PID:G325f
 A/Experimental source: strain OT3
 A/Note: this accession replaces an interim accession for a sequence replaced by GenBank
 C/Genetics:
 A/Gene: PH1819

Query Match 14.6%; Score 6; DB 2; Length 183;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 26 SKTDEA 31
 |||||
 Db 43 SKTDEA 48

Search completed: February 1, 2005, 15:33:04
 Job time : 15.5556 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:54:06 ; Search time 70.5556 Seconds
(without alignments)
334.352 Million cell updates/sec

Title: SEQ5ASP1967
Perfect score: 41
Sequence: 1 LKGLFTLPAGHLVVKFPADTL.....QVNISKTDFAFPDSENDPEK 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1825181 seqs, 575374646 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1810864

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Uniprot_02:.*
1: uniprot_sprot:.*
2: uniprot_trembl:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	349	Q96ES5	Q96ES5 homo sapien
2	41	100.0	897	Q8N7L7	Q8N7L7 homo sapien
3	41	100.0	1106	Q6P197	Q6P197 homo sapien
4	41	100.0	1106	AAH65205	AAH65205 homo sapien
5	20	48.8	349	Q8VCK1	Q8VCK1 mus musculu
6	20	48.8	408	Q8CCT5	Q8CCT5 mus musculu
7	20	48.8	958	BP28 MACPA	BP28 MACPA macaca fasc
8	20	48.8	2144	BP28 HUMAN	BP28 HUMAN homo sapien
9	8	19.5	697	Q8M9G5	Q8M9G5 homo sapien
10	7	17.1	245	Q81DM8	Q81DM8 bacillus ce
11	7	17.1	273	Q87292	Q87292 pseudomonas
12	7	17.1	296	Q9AAU6	Q9AAU6 caulobacter
13	7	17.1	366	Q8HS77	Q8HS77 corchorus c
14	7	17.1	394	Q8ROWS	Q8ROWS chitlinophag
15	7	17.1	398	Q87P73	Q87P73 vibrio para
16	7	17.1	400	Q9KSB9	Q9KSB9 vibrio chol
17	7	17.1	416	Q7WXK3	Q7WXK3 alcaligenes
18	7	17.1	422	Q9ASL8	Q9ASL8 caulobacter
19	7	17.1	453	Q9CER2	Q9CER2 lactococcus
20	7	17.1	481	Q9RB_FLEBJA	Q9RB_FLEBJA flexibacter
21	7	17.1	481	Q9AQL4	Q9AQL4 chitlinophag
22	7	17.1	481	Q9FAW9	Q9FAW9 flexibacter
23	7	17.1	481	Q9FAW9	Q9FAW9 flexibacter
24	7	17.1	663	Q9GDW7	Q9GDW7 campylobacter
25	7	17.1	705	Q8M9R0	Q8M9R0 eucommia ul
26	7	17.1	705	Q6ES570	Q6ES570 agalinia te
27	7	17.1	707	Q6ES568	Q6ES568 agalinia li
28	7	17.1	728	Q8HTM5	Q8HTM5 bogardaria c
29	7	17.1	863	Q9ADL8	Q9ADL8 polyangium
30	7	17.1	1045	HMD2_YEAST	HMD2_YEAST saccharomyc
31	7	17.1	1532	Q7RSR6	Q7RSR6 plasmidium

ALIGNMENTS

RESULT 1

32	6	14.6	64	2	Q7GY07	Q7GY07 giardia lam
33	6	14.6	67	2	Q8J373	Q8J373 human immun
34	6	14.6	67	2	Q8J381	Q8J381 human immun
35	6	14.6	67	2	Q8J382	Q8J382 human immun
36	6	14.6	67	2	Q8J383	Q8J383 human immun
37	6	14.6	73	2	Q9EM16	Q9EM16 ameba mo
38	6	14.6	75	2	Q9RVN1	Q9RVN1 deinococcus
39	6	14.6	77	2	Q90113	Q90113 human immun
40	6	14.6	80	2	Q6NGA4	Q6NGA4 corynebacte
41	6	14.6	80	2	CAES0148	CAES0148 corynebacte
42	6	14.6	87	2	Q93Y10	Q93Y10 brassica na
43	6	14.6	88	2	Q8B706	Q8B706 hepatitis c
44	6	14.6	93	2	Q6QM20	Q6QM20 bos taurus
45	6	14.6	93	2	AA548374	AA548374 bos taurus

PRELIMINARY; PRT; 349 AA.

AC Q96ES5; Q96ES5; 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE FLJ10359 protein.

GN Name=FLJ10359;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

NP [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Ovary;

RX MEDLINE=22388257; PubMed=12477932;

RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diachenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carrincci P., Prange C.,

RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,

RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,

RA Krzyzinski M.I., Skalska U., Smalins D.E., Schermer A., Schein J.E.,

RA Jones S.J., Mair M.A.,

RT "Generation and initial analysis of more than 15,000 full-length human

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Ovary;

RA Straubeberg R.L.,

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC011983; AAH11983.1; -

DR InterPro; IPR006938; ARM.

SQ SEQUENCE 349 AA; 39921 MW; 3A35597FF70799B CRC64;

Query Match 100.0%; Score 41; DB 2; Length 349;

Best Local Similarity 100.0%; Pred. No. 1,7e-34;

Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 LKGLFTLPAGHLVVKFPADTLQVNISKTDFAFPDSENDPEK 41

152 LKGLFTLPAGHLVVKFPADTLQVNISKTDFAFPDSENDPEK 192

```

RESULT 2
08N7L7 PRELIMINARY; PRT; 897 AA.
AC 08N7L7
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2003 (TREMBlrel. 22, Last sequence update)
DE Hypothetical protein FLJ40893.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RX PubMed=14702039;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Makamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuma N., Sato K., Tanikawa M., Yamazaki M.,
RA Minomiyama K., Tanishima T., Yamashita H., Murakawa M., Fujimori K.,
RA Ono Y., Takiguchi S., Watanabe M., Hirakawa S., Chiba Y., Ishida S.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Kusano J.,
RA Togiyasu K., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Arita M., Imose N.,
RA Yoshikawa Y., Matsumura H., Ichihara T., Shiohara N., Sano S.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe K., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Taniguchi H., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujitani T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirose K., Omori Y.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togoashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita K.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs."
RL Nat. Genet. 36:40-45 (2004).
DR EMBL; AK098212; BAC05261.1; -.
DR InterPro; IPR008938; ARM.
SQ SEQUENCE 897 AA; 101574 MW; 5FF6A94FE8655895 CRC64;

Query Match
Best Local Similarity 100.0%; Score 41; DB 2; Length 897;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LKGLFTLPAGHLVVPADTLTDQVNIISKTDFAFDSNDPEK 41
Db 700 LKGLFTLPAGHLVVPADTLTDQVNIISKTDFAFDSNDPEK 740

RESULT 3
06P197 PRELIMINARY; PRT; 1106 AA.
AC 06P197
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE FLJ10359 protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RX MEDLINE=22388257; PubMed=12477932;
RA Struhsberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshitsuki S., Carninci P., Prange C.,
RA Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalek U., Smalms D.E., Schnerch A., Schein J.E.,
RA Jones S.U., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA Struhsberg R.,
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC065205; AAH65205.1; -.
DR InterPro; IPR008938; ARM.
SQ SEQUENCE 1106 AA; 125359 MW; 09F7CE94042302C4 CRC64;

Query Match
Best Local Similarity 100.0%; Score 41; DB 2; Length 1106;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LKGLFTLPAGHLVVPADTLTDQVNIISKTDFAFDSNDPEK 41
Db 909 LKGLFTLPAGHLVVPADTLTDQVNIISKTDFAFDSNDPEK 949

RESULT 4
AAH65205 PRELIMINARY; PRT; 1106 AA.
AC AAH65205
DT 02-MAR-2004 (TREMBlrel. 27, Created)
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
DE FLJ10359 protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RX MEDLINE=22388257; PubMed=12477932;
RA Struhsberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshitsuki S., Carninci P., Prange C.,
RA Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

```

```

RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywnicki M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E.,
RA Jones S.J., Maira M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA Straubeberg R.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC065205; AAH6205.1; -.
FT NON TER
SO SEQUENCE 1106 AA; 125359 MW; 09F7CE94042302C4 CRC64;

Query Match 100.0%; Score 41; DB 2; Length 1106;
Best Local Similarity 100.0%; Pred. No. 4,36-34;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 LKGLFTLPAGHLVXKPFADTLDOVNSKTDPAFPSENDPEK 41
Db 909 LKGLFTLPAGHLVXKPFADTLDOVNSKTDPAFPSENDPEK 949

RESULT 5
OBVCK1 PRELIMINARY; PRT; 349 AA.
AC OBVCK1;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE BC019693 protein.
CN Name=BC019693;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=mx FVB/N;
RX TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RX MEDLINE=22388257; PubMed=12477932;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shemmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Martulina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshlyuk S., Carinci P., Prange C.,
RA Rata S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Parey J., Helton E., Kerteman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywnicki M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E.,
RA Jones S.J., Maira M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=mx FVB/N;
RX TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RA Straubeberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019693; AAH19693.1; -.
MGD; MGI:2384983; BC019693.
InterPro; IPR008938; ARM.

```

```

SQ SEQUENCE 349 AA; 40166 MW; 9763D0331AD0F515 CRC64;

Query Match 48.8%; Score 20; DB 2; Length 349;
Best Local Similarity 100.0%; Pred. No. 1,6e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 LKGLFTLPAGHLVXKPFADTL 20
Db 152 LKGLFTLPAGHLVXKPFADTL 171

RESULT 6
OBVCK5 PRELIMINARY; PRT; 408 AA.
AC OBVCK5;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 26, Last sequence update)
DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Mus musculus adult male olfactory brain cDNA, RIKEN full-length
DE enriched library, clone:643040D06 product:hypothetical ARM repeat
DE structure containing protein, full insert sequence.
CN Name=BC019693;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=99279253; PubMed=10349636;
RA Carinci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning."
RL Mech. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20499374; PubMed=11042159;
RA Carinci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Komo H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes."
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carinci P.,
RA Kono H., Akiyama J., Nishi K., Kitana T., Tashiro H., Itoh M.,
RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kaishiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanuki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer."
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.

```

RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoaka T., Hirozane T.,
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
 RA Kato H., Kawai J., Kojima Y., Kondo S., Kono H., Koyama S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takehashi F., Takaku-Akahara S., Takeda Y., Tanaka T.,
 RA Tomaru A., Taya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
 RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK032134; BAC27721.1; -
 DR MGI; 2384983; BC019693.
 DR InterPro; IPR008938; ARM.
 KW Hypothetical protein.
 SQ SEQUENCE 408 AA; 4668 MW; 42AB9EB13CA3FE67 CRC64;

Query Match 48.8%; Score 20; DB 2; Length 408;
 Best Local Similarity 100.0%; Pred. No. 1.9e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 LKGLFTLPAGHLVXKPPADTL 20
 Db 211 LKGLFTLPAGHLVXKPPADTL 230

RESULT 7
 BP28_MACFA STANDARD; PRT; 958 AA.
 AC O9GM24;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Protein BAP28 (OmpA-17571) (Fragment).
 GN Name=BAP28;
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 CC Cercopithecinae; Macaca.
 OK NCBI_TaxID=9541;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC Tissue=Brain;
 RA Osada N., Hida M., Kusuda J., Tanuma R., Ieiki K., Hirai M., Terao K.,
 RA Suzuki Y., Sugano S., Hashimoto K.;
 RT "Isolation of full-length cDNA clones from macaque brain cDNA
 RT libraries.";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 CC -1 SIMILARITY: Belongs to the BAP28 family.
 CC -1 SIMILARITY: Contains 1 HEAT repeat.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB049842; BAB16728.1; ALT_INIT.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR000357; HEAT.
 DR PROSITE; PS50077; HEAT_REPEAT; FALSE_NEG.
 FT NON_TER 1
 FT REPEAT 1
 SQ SEQUENCE 958 AA; 108644 MW; 3DBD95C6323CFB31 CRC64;

Query Match 48.8%; Score 20; DB 1; Length 958;
 Best Local Similarity 100.0%; Pred. No. 3.8e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LKGLFTLPAGHLVXKPPADTL 20
 Db 761 LKGLFTLPAGHLVXKPPADTL 780

RESULT 8
 BP28_HUMAN STANDARD; PRT; 2144 AA.
 AC O9H583; O9NM23;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Protein BAP28.
 GN Name=BAP28;
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OK NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A. AND VARIANTS SER-1694; ALA-1854; ASP-1967 AND
 RP GLY-2017.
 RA Bouguetel L., Chumakov I., Barry C., Cohen-Akenine A.;
 RT "A novel BAP28 gene and protein."
 RL Patent number WO0100669, 04-JAN-2001.
 RN (2)
 RP SEQUENCE OF 1534-2144 FROM N.A.
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE OF 1777-2144 FROM N.A.
 RX PubMed=14702039; DOI=10.1038/ng1285;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
 RA Nagatani A., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
 RA Shiratori A., Sudo H., Hosoliri T., Kaku Y., Kodaira H., Kondo H.,
 RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
 RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
 RA Yamazaki M., Ninomiya K., Ihibashi T., Yamashita H., Murakawa K.,
 RA Fujimori K., Tanai H., Kimura M., Watanabe M., Hirao K., Chiba Y.,
 RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hottu T.,
 RA Kusano Y., Taniguchi K., Takahashi F., Hara R., Takeuchi K., Arita M.,
 RA Inoue N., Mitsuhashi K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Omori Y.,
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senda T.,
 RA Matsunuma K., Nakajima Y., Mizuno T., Morinaga M., Komatsu T.,
 RA Togauchi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Sato T., Shirai Y., Takahashi K., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Iio T., Sugano S.;
 RT "Complete sequencing and characterization of 21,243 full-length human
 RT cDNAs.";
 RL Nat. Genet. 36:40-45(2004).
 CC -1 SIMILARITY: Belongs to the BAP28 family.
 CC -1 SIMILARITY: Contains 1 HEAT repeat.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

DR EMBL; AX067150; CAC26776.1; -.
DR EMBL; AL36105; CAC15948.1; -.
DR EMBL; AK011221; BAA91564.1; ALT_INIT.
DR SWISS-2DPAGE; Q9H583; HUMAN.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000357; HEAT.
DR PROSITE; PS50077; HEAT_REPEAT; FALSE_NEG.
KM Polymorphism.
FT REPEAT 2106 2142 HEAT.
FT VARIANT 1694 1694 N->S.
FT VARIANT 1854 1854 /FTId=VAR_010939.
FT VARIANT 1967 1967 V->A.
FT VARIANT 1967 1967 /FTId=VAR_010940.
FT VARIANT 2017 2017 N->D.
FT VARIANT 2017 2017 /FTId=VAR_010941.
SQ SEQUENCE 2144 AA; 242355 MW; D66816EE78DC9B7 CRC64;
Query Match 48.8%; Score 20; DB 1; Length 2144;
Best Local Similarity 100.0%; Pred. No. 7,3e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LKGLFTLPAGHLVVPADTL 20
Db 1947 LKGLFTLPAGHLVVPADTL 1966

RESULT 9
Q8M9G5 PRELIMINARY; PRT; 697 AA.
ID Q8M9G5;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE NMDH dehydrogenase subunit F (Fragment).
GN Name=ndhF;
OS Kaliphora madagascariensis.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Solanales; Montiniaceae; Kaliphora.
OC NCBI_TaxID=125030;
RN [1]
RP SEQUENCE FROM N.A.
RA Bremer B., Bremer K., Heidari N., Erixon P., Olmstead R.G.,
RA Anderberg A.A., Kallersjö M., Barkhoradian E.;
RT "Phylogenetics of asterids based on 3 coding and 3 non-coding
RT chloroplast DNA markers and the utility of non-coding DNA at higher
RT taxonomic levels.";
RL Mol. Phylogenet. Evol. 24:273-300(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Lundberg J.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ431206; CAD24024.1; -.
DR GO; GO:0009507; C:chloroplast; IEA.
DR GO; GO:0009553; C:photosystem II; IEA.
DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0042773; P:ATP synthesis coupled electron transport; IEA.
DR InterPro; IPR003945; NADH1_oxred5.
DR InterPro; IPR003916; NADH1b_oxred5.
DR InterPro; IPR001750; Oxidored_q1.
DR InterPro; IPR00128; Oxidored_q1_C.
DR InterPro; IPR001516; Oxidored_q1_N.
DR Pfam; PF00361; Oxidored_q1; 1.
DR Pfam; PF01010; Oxidored_q1_C; 1.
DR Pfam; PF00662; Oxidored_q1_N; 1.
DR PRINTS; PR01434; NADH1b_NADH1b.
DR PRINTS; PR01435; NADH1b_NADH1b.
DR TIGRfam; TIGR01974; Ndh_1_L; 1.
KM Chloroplast; NAD; NADP; Oxidoreductase; Plastoguinone; Quinone.

```

```

FT NON TER 1 1
FT NON TER 697 697
SQ SEQUENCE 697 AA; 78927 MW; 27E1B753AB164423 CRC64;
Query Match 19.5%; Score 8; DB 2; Length 697;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GLFTLPAG 10
Db 547 GLFTLPAG 554

RESULT 10
Q81DM8 PRELIMINARY; PRT; 245 AA.
ID Q81DM8;
AC Q81DM8;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Zinc uptake transporter.
GN ORFNames=BC2329;
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OC NCBI_TaxID=226900;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630; DOI=10.1038/nature01582;
RA Ivanova N., Sorokin A., Anderson I., Gallerton N., Candelion B.,
RA Kapatarel V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Coleman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Fusch G., Haselkorn R., Fomstein M., Ehrlich S.D.,
RA Overbeek R., Kyprides N.C.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis.";
RL Nature 423:87-91(2003).
DR EMBL; AE017005; AAP09293.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0046873; F:metal ion transporter activity; IEA.
DR GO; GO:0030001; P:metal ion transport; IEA.
DR InterPro; IPR003689; Zn_transp_Zip.
DR Pfam; PF02535; Zip; 1.
SQ SEQUENCE 245 AA; 27087 MW; EF5C8D3CEFB9C934 CRC64;
Query Match 17.1%; Score 7; DB 2; Length 245;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 FTLPAGH 11
Db 238 FTLPAGH 244

RESULT 11
Q87Z92 PRELIMINARY; PRT; 273 AA.
ID Q87Z92;
AC Q87Z92;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Bacterial transferase, hexapeptide repeat protein.
GN OrderedLocustNames=PSPT03538;
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OC NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;
RA Bueli C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,
RA Gwin M.L., Dodson R.J., DeBoy R.T., Durkin A.S., Kollonay J.F.,
RA Madupu R., Daugherty S.C., Brinkac L.M., Beanan M.J., Haft D.H.,

```

RA Nelson W.C., Davidsen T.M., Zafar N., Zhou L., Liu J., Yuan Q.,
 RA Khouri H.M., Fedorova N.B., Tran B., Russell D., Berry K.J.,
 RA Uterback T.R., Van Aken S.E., Feldblum T.V., D'Ascenzo M.,
 RA Deng W.-L., Ramos A.R., Alfano J.R., Gartinour S., Chatterjee A.K.,
 RA Belaney I.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,
 RA Bender C.L., White O., Fraser C.M., Collier A.,
 RT "The complete genome sequence of the Arabidopsis and tomato pathogen
 Pseudomonas syringae pv. tomato DC3000.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).
 DR EMBL; AF016868; AAC57013.1; -.
 DR TIGR; PSP03538; -.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR InterPro; IPR001451; Hexasep_transf.
 DR InterPro; IPR011004; Trimer_LpxA_like.
 DR Pfam; PF00132; Hexapep; 4.
 KW Complete proteome; Transferase.
 SQ SEQUENCE 273 AA; 30490 MW; 6005A00571620D12 CRC64;

Query Match 17.1%; Score 7; DB 2; Length 273;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KGLFTLF 8
 Db 266 KGLFTLF 272

RESULT 12

Q9AUA6 PRELIMINARY; PRT; 296 AA.
 AC Q9AUA6; -.
 DT 01-JUN-2001 (TRENBLrel. 17; Created)
 DT 01-JUN-2001 (TRENBLrel. 17; Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26; Last annotation update)
 DE Hypothetical protein CC0498.
 GN OrderedLocustName=CC0498;
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
 OC Caulobacteraceae; Caulobacter.
 OX NCBI_TaxID=155892;
 RN [1]

RP SEQUENCE FROM N.A. / CB15;
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647; DOI=10.1073/pnas.061029298;
 RA Nieman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Knouri H.M., Shetty J.,
 RA Berry K.J., Uterback T.R., Tran K., Wolf A.M., Vamathavan J.J.,
 RA Ermolaeva M.D., White O., Salzberg S.L., Venter J.C., Shapiro L.,
 RA Fraser C.M.;
 RT "Complete genome sequence of Caulobacter crescentus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL; AE005722; AAC2485.1; -.
 DR PIR; A87311; A87311.
 DR TIGR; CC0498; -.
 DR InterPro; IPR003169; GYF.
 DR InterPro; IPR010432; RPD.
 DR Pfam; PF06271; RPD; 1.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 296 AA; 32415 MW; B9D92D2D8A921156 CRC64;

Query Match 17.1%; Score 7; DB 2; Length 296;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 LFTLFLAG 10
 Db 240 LFTLFLAG 246

RESULT 13

Q8H577 PRELIMINARY; PRT; 366 AA.
 ID Q8H577
 AC Q8H577;
 DT 01-MAR-2003 (TRENBLrel. 23; Created)
 DT 01-MAR-2003 (TRENBLrel. 23; Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25; Last annotation update)
 DE NADH dehydrogenase (Fragment).
 GN Name=ndhF;
 OS Corchorus capsularis.
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; rosids;
 OC eurosids II; Malvales; Malvaceae; Grewioideae; Corchorus.
 OX NCBI_TaxID=210143;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Whitlock B.A., Karol K.G., Alverson W.S.;
 RT "Chloroplast DNA sequences confirm the placement of the enigmatic
 RT Oceanopavaver within Corchorus (Grewioideae: Malvaceae s.l., formerly
 RT Tiliaceae).";
 RL Int. J. Plant Sci. 164:35-41(2003).
 DR EMBL; AF523838; AA016032.1; -.
 DR GO; GO:0009507; C:chloroplast; IEA.
 DR GO; GO:0009523; C:photosystem II; IEA.
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0042773; P:ATP synthesis coupled electron transport; IEA.
 DR InterPro; IPR003945; NADHl_oxrds.
 DR InterPro; IPR001750; Oxidored_q1.
 DR InterPro; IPR002128; Oxidored_q1_C.
 DR Pfam; PF00361; Oxidored_q1; 1.
 DR Pfam; PF01010; Oxidored_q1_C; 1.
 DR PRINTS; PR01435; NPOXDRDTASE5.
 KW Chloroplast; NAD; NADP; Oxidoreductase; Plastocyanine; Quinone.
 FT NON_TER 1 1
 FT TER 366 366
 SQ SEQUENCE 366 AA; 41625 MW; A6E391ABF5CA2A6A CRC64;

Query Match 17.1%; Score 7; DB 2; Length 366;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 LFTLFLAG 10
 Db 229 LFTLFLAG 235

RESULT 14

Q8R0W5 PRELIMINARY; PRT; 394 AA.
 ID Q8R0W5
 AC Q8R0W5;
 DT 01-JUN-2002 (TRENBLrel. 21; Created)
 DT 01-JUN-2002 (TRENBLrel. 21; Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25; Last annotation update)
 DE DNA gyrase subunit B (EC 5.99.1.3) (Fragment).
 GN Name=gyrB;
 OS Chitinophaga pinensis.
 OC Bacteria; Bacteroidetes; Sphingobacteriales; Sphingobacteriales;
 OC Crenotrichaceae; Chitinophaga.
 OX NCBI_TaxID=79329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM2588;
 RX MEDLINE=22413356; PubMed=12542710;
 RA Matsuno Y., Suzuki M., Kasai H., Shizuri Y., Haryama S.;
 RT "Isolation and phylogenetic characterization of bacteria capable of
 RT inducing differentiation in the green alga Monostroma oxyspermum.";
 RL Environ. Microbiol. 5:25-35(2003).
 CC -1- FUNCTION: DNA gyrase negatively supercoils closed circular double-
 CC stranded DNA in an ATP-dependent manner and also catalyzes the
 CC interconversion of other topological isomers of double-stranded
 CC DNA rings, including catenanes and knotted rings (by similarity).
 CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining

CC of double-stranded DNA.
 CC -1- SIMILARITY: Belongs to the type II topoisomerase family.
 DR EMBL: AB073077; BAB88379.1; -.
 DR HSSP: P06982; 1AJ6
 DR GO: GO:0005524; F:ATP binding; IEA.
 DR GO: GO:0003677; F:DNA binding; IEA.
 DR GO: GO:0003918; F:DNA topoisomerase (ATP-hydrolyzing) activity; IEA.
 DR GO: GO:0016853; F:isomerase activity; IEA.
 DR GO: GO:0006265; P:DNA topological change; IEA.
 DR InterPro: IPR003594; ATPbind_ATPase.
 DR InterPro: IPR01241; DNA_topoisomII.
 DR Pfam: PF00204; DNA_gyraseb; 1.
 DR Pfam: PF02518; HATPase_C; 1.
 DR PRINTS: PR00418; TP12FAMILY.
 DR SMART: SM00433; TOP2C; 1.
 DR PROSITE: PS00177; TOPOISOMERASE_II; 1.
 DR ATP-binding; Isomerase; Topoisomerase.
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 394 AA; 43691 MW; 18A2C158BD54AABF CRC64;

Query March 17.1%; Score 7; DB 2; Length 394;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 SENDPEK 41
 |||||
 Db 310 SENDPEK 316

RESULT 15

087P73 PRELIMINARY; PRT; 398 AA.
 AC 087P73;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein VP1645.
 GN OrderedLocustName=VP1645;
 OS *Vibrio parahaemolyticus*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; *Vibrio*.
 OC NCBI_TaxID=670;
 OX 11
 RP SEQUENCE FROM N. A.
 RC STRAIN=RTMD 2210633 / Serotype O3:K6;
 RX MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Najiima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yaunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
 RT "Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism
 distinct from that of *V. cholerae*.";
 RL Lancet 361:743-749 (2003).
 DR EMBL: AP005078; BAC59908.1; -.
 DR InterPro: IPR007400; DUF453.
 DR Pfam: PF04303; DUF453; 1.
 KM Complete proteome; Hypothetical protein
 SQ SEQUENCE 398 AA; 42358 MW; 8CPE4C76CED989B0 CRC64;

Query March 17.1%; Score 7; DB 2; Length 398;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 QVNISKI 28
 |||||
 Db 143 QVNISKI 149

Search completed: February 1, 2005, 15:31:04
 Job time : 72.5536 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 15:09:39 ; Search time 59.7778 Seconds
(without alignments)
247.799 Million cell updates/sec

Title: SEQ5GLU2017

Perfect score: 41
Sequence: 1 NCLYKIFLPTQHFISKERA.....ALMPLVDLENRLGGEK 41

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 1608061 seqs, 361289386 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1311380

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US05_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
- 19: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	1149	17	US-10-128-558-167
2	8	19.5	378	15	US-10-282-122A-49544
3	7	17.1	53	15	US-10-424-599-275868
4	7	17.1	238	16	US-10-767-701-41775
5	7	17.1	348	14	US-10-369-493-28801
6	7	17.1	349	15	US-10-424-599-149524
7	7	17.1	406	15	US-10-425-114-53502
8	7	17.1	633	16	US-10-437-963-110339
9	7	17.1	660	16	US-10-437-963-151847
10	7	17.1	662	16	US-10-437-963-110342
11	7	17.1	665	15	US-10-424-599-149522
12	6	14.6	36	9	US-09-864-761-44493
13	6	14.6	45	17	US-10-425-115-340619

14	6	14.6	48	9	US-09-989-920-197	Sequence 197, App
15	6	14.6	49	17	US-10-425-115-339911	Sequence 339911,
16	6	14.6	51	15	US-10-424-599-219715	Sequence 219715,
17	6	14.6	52	14	US-10-029-386-29247	Sequence 29247, A
18	6	14.6	55	17	US-10-425-115-203297	Sequence 203297,
19	6	14.6	55	15	US-10-424-599-245649	Sequence 245649,
20	6	14.6	65	9	US-09-815-242-12049	Sequence 12049, A
21	6	14.6	65	14	US-10-127-032-135	Sequence 135, App
22	6	14.6	65	15	US-10-282-122A-66719	Sequence 66719, A
23	6	14.6	65	15	US-10-389-647-644	Sequence 644, App
24	6	14.6	72	16	US-10-437-963-178932	Sequence 178932,
25	6	14.6	73	17	US-10-425-115-266081	Sequence 266081,
26	6	14.6	74	17	US-10-425-115-364110	Sequence 364110,
27	6	14.6	87	17	US-10-425-115-295263	Sequence 295263,
28	6	14.6	87	16	US-10-767-701-54325	Sequence 54325, A
29	6	14.6	97	16	US-10-437-963-180730	Sequence 180730,
30	6	14.6	100	16	US-10-437-963-137984	Sequence 137984,
31	6	14.6	102	15	US-10-282-122A-52746	Sequence 52746, A
32	6	14.6	102	15	US-10-627-476-94	Sequence 94, Appl
33	6	14.6	102	16	US-10-767-701-54628	Sequence 54628, A
34	6	14.6	103	15	US-10-424-599-205861	Sequence 205861,
35	6	14.6	107	17	US-10-425-115-333372	Sequence 333372,
36	6	14.6	112	17	US-10-425-115-219369	Sequence 219369,
37	6	14.6	118	16	US-10-437-963-148111	Sequence 148111,
38	6	14.6	119	17	US-10-425-114-39348	Sequence 259301,
39	6	14.6	130	15	US-10-425-114-39348	Sequence 39348, A
40	6	14.6	132	15	US-10-424-599-262325	Sequence 262325,
41	6	14.6	138	15	US-10-424-599-280443	Sequence 280443,
42	6	14.6	140	15	US-10-425-114-41418	Sequence 41418, A
43	6	14.6	141	14	US-10-117-087-4	Sequence 4, Appl1
44	6	14.6	141	14	US-10-117-087-6	Sequence 6, Appl1
45	6	14.6	141	16	US-10-437-963-102617	Sequence 102617,

ALIGNMENTS

RESULT 1
US-10-128-558-167
Sequence 167, Application US/10128558
Publication No. US2004021952A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Wang, Zhiwei
APPLICANT: Weng, Gezhi
APPLICANT: Boyle, Bryan J
APPLICANT: Drmanac, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
FILE REFERENCE: 812A
CURRENT APPLICATION NUMBER: US/10/128, 558
CURRENT FILING DATE: 2002-04-22
PRIOR APPLICATION NUMBER: US 60/339,453
PRIOR FILING DATE: 2001-12-11
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02823
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: PCT/US01/03800
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 09/515,126
PRIOR FILING DATE: 2000-02-28
Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 412
 SOFTWARE: pt_FL_genes Version 6.0
 SEQ ID NO: 167
 LENGTH: 1149
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-128-558-167

Query Match 100.0%; Score 41; DB 17; Length 1149;
 Best Local Similarity 100.0%; Pred. No. 1.8e-33;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NCLYKIFLFDTHFISKERAALMPLVQLNRLGGEKEF 41
 DB 1002 NCLYKIFLFDTHFISKERAALMPLVQLNRLGGEKEF 1042

RESULT 2
 US-10-282-122A-49544
 Sequence 49544, Application US/10282122A
 Publication No. US20040029129A1
 GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
 APPLICANT: Zamudio, Carlos
 APPLICANT: Malone, Cheryl
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Kari
 APPLICANT: Zyskind, Judith
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John
 APPLICANT: Carr, Grant
 APPLICANT: Yamamoto, Robert
 APPLICANT: Foreyth, R.
 APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 FILE REFERENCE: ELITRA.034A
 CURRENT APPLICATION NUMBER: US/10/282,122A

PRIOR FILING DATE: 2003-02-20
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/230,335
 PRIOR FILING DATE: 2000-09-06
 PRIOR APPLICATION NUMBER: 60/230,347
 PRIOR FILING DATE: 2000-09-09
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/267,636
 PRIOR FILING DATE: 2001-02-09
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 Remaining Prior Application data removed - See File Wrapper or PALM.
 NUMBER OF SEQ ID NOS: 78614
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO: 49544
 LENGTH: 378
 TYPE: PRT
 ORGANISM: Burkholderia fungorum
 US-10-282-122A-49544

Query Match 19.5%; Score 8; DB 15; Length 378;
 Best Local Similarity 100.0%; Pred. No. 6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 SKERAAL 23
 |||||

DB 310 SKERAAL 317

RESULT 3
 US-10-424-599-275868
 Sequence 275868, Application US/10424599
 Publication No. US20040031072A1
 GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J
 APPLICANT: Kovalic, David K
 APPLICANT: Zhou, Yihua
 APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 FILE REFERENCE: 38-21(53223)B
 CURRENT APPLICATION NUMBER: US/10/424,599

PRIOR FILING DATE: 2003-04-28
 NUMBER OF SEQ ID NOS: 285684
 SEQ ID NO: 275868
 LENGTH: 53
 TYPE: PRT
 ORGANISM: Glycine max

OTHER INFORMATION: Clone ID: PAT_MRT3847_91128C.1.pep
 US-10-424-599-275868

Query Match 17.1%; Score 7; DB 15; Length 53;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 DTQHFIS 16
 DB 32 DTQHFIS 38
 |||||

RESULT 4
 US-10-767-701-41775

Sequence 41775, Application US/10767701
 Publication No. US20040172684A1
 GENERAL INFORMATION:

APPLICANT: Kovalic, David K.
 APPLICANT: Zhou, Yihua
 APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 FILE REFERENCE: 38-21(53535)B
 CURRENT APPLICATION NUMBER: US/10/767,701

PRIOR FILING DATE: 2004-01-29
 NUMBER OF SEQ ID NOS: 63128
 SEQ ID NO: 41775
 LENGTH: 238
 TYPE: PRT
 ORGANISM: Sorghum bicolor

OTHER INFORMATION: Clone ID: SORBI-28MAY03-C6863_1.pep
 US-10-767-701-41775

Query Match 17.1%; Score 7; DB 16; Length 238;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 LGGEKEF 41
 DB 211 LGGEKEF 217
 |||||

RESULT 5
 US-10-369-493-22801

Sequence 22801, Application US/10369493
 Publication No. US20030233675A1
 GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.

```

; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22801
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-22801

Query Match      17.1% Score 7; DB 14; Length 348;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      17 KERAAL 23
        |||||
        141 KERAAL 147

Db

RESULT 6
US-10-424-599-149524
; Sequence 149524, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 149524
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(349)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_106042C.1.pdp
US-10-424-599-149524

Query Match      17.1% Score 7; DB 15; Length 349;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      26 PLVDQLE 32
        |||||
        290 PLVDQLE 296

Db

RESULT 7
US-10-425-114-53502
; Sequence 53502, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
```

```

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53113)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 53502
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 701129526_FLI.pdp
US-10-425-114-53502

Query Match      17.1% Score 7; DB 15; Length 406;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      26 PLVDQLE 32
        |||||
        335 PLVDQLE 341

Db

RESULT 8
US-10-437-963-110339
; Sequence 110339, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 110339
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_14411C.1.pdp
US-10-437-963-110339

Query Match      17.1% Score 7; DB 16; Length 633;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      26 PLVDQLE 32
        |||||
        559 PLVDQLE 565

Db

RESULT 9
US-10-437-963-151847
; Sequence 151847, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
```

```

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 151847
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_51953C.1.pep
US-10-437-963-151847
```

```

Query Match      17.1%; Score 7; DB 16; Length 660;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      26 PLVDQLE 32
      |||||
Db      588 PLVDQLE 594
```

```

RESULT 10
US-10-437-963-110342
; Sequence 110342, Application US/10437963
; Publication No. US2004012333A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 110342
; LENGTH: 662
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_14414C.1.pep
US-10-437-963-110342
```

```

Query Match      17.1%; Score 7; DB 16; Length 662;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      26 PLVDQLE 32
      |||||
Db      588 PLVDQLE 594
```

```

RESULT 11
US-10-424-599-149522
; Sequence 149522, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
```

```

; SEQ ID NO 149522
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(665)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_106040C.1.pep
US-10-424-599-149522
```

```

Query Match      17.1%; Score 7; DB 15; Length 665;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      26 PLVDQLE 32
      |||||
Db      594 PLVDQLE 600
```

```

RESULT 12
US-09-864-761-44493
; Sequence 44493, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
```

```
SEQ ID NO 44493
LENGTH: 36
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005229.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.96
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.9
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.86
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.86
US-09-864-761-44493

Query Match
Best Local Similarity 100.0%; Score 6; DB 9; Length 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 GGEKFF 41
DB 4 GGEKFF 9

RESULT 13
US-10-425-115-340619
Sequence 340619, Application US/10425115
Publication No. US20040214272A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants
FILE REFERENCE: 38-21(53222)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 340619
LENGTH: 45
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_73815C.1.pcp
US-10-425-115-340619

Query Match
Best Local Similarity 100.0%; Score 6; DB 17; Length 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 LYKIFL 8
DB 14 LYKIFL 19

RESULT 14
US-09-989-920-197
Sequence 197, Application US/09989920
Patent No. US20020172957A1
GENERAL INFORMATION:
APPLICANT: Macina, Roberto
APPLICANT: Recipon, Herve
APPLICANT: Chen, Sel-Yu
APPLICANT: Sun, Yongming
APPLICANT: Liu, Chenghua
TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prot
FILE REFERENCE: DEX-0291
CURRENT APPLICATION NUMBER: US/09/989,920
CURRENT FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: 60/252,500
NUMBER OF SEQ ID NOS: 284
SOFTWARE: PatentIn version 3.1
```

```
SEQ ID NO 197
LENGTH: 48
TYPE: PRT
ORGANISM: Homo sapien
US-09-989-920-197

Query Match
Best Local Similarity 100.0%; Score 6; DB 9; Length 48;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 PLVDQL 31
DB 42 PLVDQL 47

RESULT 15
US-10-425-115-339911
Sequence 339911, Application US/10425115
Publication No. US20040214272A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants
FILE REFERENCE: 38-21(53222)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 339911
LENGTH: 49
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
NAME/KEY: unsure
LOCATION: (1) ..(49)
OTHER INFORMATION: unsure at all Xaa locations
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_73167C.1.pcp
US-10-425-115-339911

Query Match
Best Local Similarity 100.0%; Score 6; DB 17; Length 49;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 RAERLM 24
DB 21 RAERLM 26
```

Search completed: February 1, 2005, 15:44:52
Job time : 59.7778 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:55:01 ; Search time 13.5556 Seconds
(without alignments)
291.016 Million cell updates/sec

Title: SEQ5GLU2017
Perfect score: 41
Sequence: 1 NCLYKIFLFDYQHFKSKERA.....ALMMPLVQLENRLGGEKRF 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 279530

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	17.1	232	2 JC7706	beta crystallin B1
2	7	17.1	276	2 AG3305	23S ribosomal RNA
3	7	17.1	362	2 T41163	cholephosphatase c
4	7	17.1	396	2 S38924	hypothetical prote
5	7	17.1	584	2 S77647	mobilisation prote
6	7	17.1	651	2 T06979	polyadenylate-bind
7	7	17.1	651	2 A21851	22K factor - human
8	7	17.1	65	2 F83054	conserved hypochet
9	7	17.1	132	1 I40369	H-transferring tw
10	7	17.1	154	2 D39384	finger protein HTP
11	7	17.1	165	2 E70385	conserved hypochet
12	7	17.1	184	2 AE1538	transcription regu
13	7	17.1	188	2 T21208	hypothetical prote
14	7	17.1	196	2 G65039	hypothetical prote
15	7	17.1	196	2 F90919	repressor for uid
16	7	17.1	196	2 C85768	glucuronide repres
17	7	17.1	196	2 D64918	conserved hypochet
18	7	17.1	204	2 F87295	conserved hypochet
19	7	17.1	211	2 E83379	conserved hypochet
20	7	17.1	229	2 T08040	MADS-box protein -
21	7	17.1	232	2 H69173	conserved hypochet
22	7	17.1	235	2 B41845	orf B - Treponema
23	7	17.1	239	2 AE0883	phosphoproteinase
24	7	17.1	241	1 RBN218	phosphoproteinase
25	7	17.1	245	2 E90239	tryptophan synhas
26	7	17.1	247	2 T08455	hypothetical prote
27	7	17.1	249	2 E86231	hypothetical prote
28	7	17.1	249	2 S30580	U2 snRNP protein A
29	7	17.1	250	2 T26010	hypothetical prote

30	6	14.6	251	2 C84036	succinate dehydrog
31	6	14.6	251	2 C83083	conserved hypochet
32	6	14.6	256	2 E72257	hypothetical prote
33	6	14.6	264	2 G83165	probable Arp-bindi
34	6	14.6	268	2 T02448	hypothetical prote
35	6	14.6	269	2 AD2833	GAD67 family prote
36	6	14.6	270	2 D90542	conserved hypochet
37	6	14.6	281	2 T43945	ribosomal protein
38	6	14.6	283	2 H97610	hypothetical prote
39	6	14.6	291	2 G97327	short-chain alcoho
40	6	14.6	292	2 B41556	clindamycin resist
41	6	14.6	295	2 D90252	conserved hypochet
42	6	14.6	302	2 T08522	NTP-binding protei
43	6	14.6	302	2 S32178	NTP-binding protei
44	6	14.6	302	2 S70151	tnc protein homol
45	6	14.6	315	2 C87293	hypothetical prote

ALIGNMENTS

RESULT 1

JC7706
beta crystallin B1 protein - zebra fish
C/Species: Brachydanio rerio (zebra fish)
C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Jul-2004
C/Accession: JC7706

R/Chen, J.Y.; Chang, B.E.; Chen, Y.H.; Lin, C.J.F.; Wu, J.L.; Kuo, C.M.

Biochem. Biophys. Res. Commun. 285, 105-110, 2001

A/Title: Molecular cloning, developmental expression, and hormonal regulation of zebrafish

A/Reference number: JC7706; MUID:21331298; PMID:11437379

A/Accession: JC7706

A/Molecule type: mRNA

A/Residues: 1-232 <CHE>

A/Cross-references: UNIPROT:Q90WT1; GB:A317957

C/Comment: This protein is involved in regulation by growth factors.

C/Superfamily: beta-crystallin

F/20-25/Region: active proline- and alanine-rich motif #status predicted

Query Match 17.1%; Score 7; DB 2; Length 232;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 YKIFLFD 10
DB 41 YKIFLFD 47

RESULT 2

AG3305
23S ribosomal RNA methyltransferase (BC 2.1.1.-) [imported] - Brucella melitensis (strain
C/Species: Brucella melitensis
C/Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C/Accession: AG3305

R/DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, I

; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Leveser

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A/Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis.

A/Reference number: AD3252; PMID:11756688

A/Accession: AG3305

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-276 <KIR>

A/Cross-references: UNIPROT:Q8Y1L5; UNIPROT:Q8F2A1; GB:AE008917; PIDD:AA151610.1; PID:gl

A/Experimental source: strain 16M

C/Genetics:

A/Gene: BMEI0429

A/Map position: 1

C/Keywords: methyltransferase

Query Match 17.1%; Score 7; DB 2; Length 276;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 LENRLGG 37
|||||
Db 67 LENRLGG 73

RESULT 3

T41163
cholelinophosphate cytidyllyltransferase - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C/Accession: T41163
R/Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, June 1998
A/Reference number: Z21974
A/Accession: T41163
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-362 <SEE>
A/Cross-references: UNIPROT:O74975; EMBL:AL023777; PIDN:CAA19310.1; GSPDB:GN00068; SPDB:
A/Experimental source: strain 972h-; cosmid c1827
C/Genetics:
A/Map position: 3
C/Superfamily: choline-phosphate cytidyllyltransferase

Query Match 17.1%; Score 7; DB 2; Length 362;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 KERAEAL 23
|||||
Db 155 KERAEAL 161

RESULT 4

S38924
hypochemical protein 12 - phage phi-C31
C/Species: phage phi-C31
C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jul-2004
C/Accession: S38924
R/Hartley, N.M.; Murphy, G.O.; Bruton, C.J.; Chater, K.F.
submitted to the EMBL Data Library, November 1993
A/Reference number: S38912
A/Accession: S38924
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-396 <HMR>
A/Cross-references: UNIPROT:Q38033; EMBL:X76288; NID:G432610; PIDN:CAA53908.1; PID:G5790
C/Genetics:
A/Start codon: GTG

Query Match 17.1%; Score 7; DB 2; Length 396;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 SKERAEA 22
|||||
Db 146 SKERAEA 152

RESULT 5

S77647
mobilisation protein A - Lactococcus lactis self-transfer sex factor
C/Species: Lactococcus lactis
C/Date: 24-Oct-1998 #sequence_revision 24-Oct-1998 #text_change 21-Jul-2000
C/Accession: S77647
R/Shearman, C.A.; Godon, J.J.; Gasson, M.
Mol. Microbiol. 21, 45-53, 1996
A/Title: Splicing of a group II intron in a functional transfer gene of Lactococcus lact
A/Reference number: S77646; MUID:97000348; PMID:8843453
A/Accession: S77647
A/Molecule type: DNA

A/Residues: 1-584 <SHR>
A/Cross-references: EMBL:X89922; NID:G1296826; PIDN:CAA61995.1; PID:G1296828
A/Experimental source: strain MG1363
C/Genetics:
A/Gene: mobA
A/Mobile element: self-transfer sex factor
A/Introns: 171/3
C/Function:
A/Description: involved in conjugation, probably by introducing a single-stranded nick at
C/Superfamily: Lactococcus lactis self-transfer sex factor mobilisation protein A
C/Keywords: conjugation; replication

Query Match 17.1%; Score 7; DB 2; Length 584;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 VDQLENR 34
|||||
Db 396 VDQLENR 402

RESULT 6

T06979
polyadenylate-binding protein - wheat
N/Alternate names: poly(A)-binding protein
C/Species: Triticum aestivum (common wheat)
C/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004
C/Accession: T06979
R/Le, H.; Chang, S.C.; Tanguay, R.L.; Gallie, D.R.
Eur. J. Biochem. 243, 350-357, 1997
A/Title: The wheat poly(A)-binding protein functionally complements pabl in yeast.
A/Reference number: Z12044; MUID:97182620; PMID:9930759
A/Accession: T06979
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-651 <LEH>
A/Cross-references: UNIPROT:P93616; EMBL:U81318; NID:G1737491; PIDN:AAB38974.1; PID:G173
A/Experimental source: sprout tips from 5 day old growing sprouts
C/Genetics:
A/Note: wheatpab

C/Superfamily: polyadenylate-binding protein; ribonucleoprotein repeat homology
C/Keywords: nucleus; RNA binding
F/33-100/Domain: ribonucleoprotein repeat homology <RRM1>
F/121-187/Domain: ribonucleoprotein repeat homology <RRM2>
F/212-278/Domain: ribonucleoprotein repeat homology <RRM3>
F/315-381/Domain: ribonucleoprotein repeat homology <RRM4>

Query Match 17.1%; Score 7; DB 2; Length 651;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 PLVDQLE 32
|||||
Db 578 PLVDQLE 584

RESULT 7

A21851
22K factor - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 28-Sep-1987 #sequence_revision 28-Sep-1987 #text_change 09-Jul-2004
C/Accession: A21851
R/Damme, J.V.; De Ley, M.; Opdenaker, G.; Billiau, A.; De Somer, P.
Nature 314, 266-268, 1985
A/Title: Homogeneous interferon-inducing 22K factor is related to endogenous pyrogen and
A/Reference number: A21851; MUID:85163727; PMID:3920526
A/Accession: A21851
A/Status: preliminary
A/Molecule type: protein
A/Residues: 1-38 <DAM>
A/Cross-references: UNIPROT:O7M4S7
C/Superfamily: interleukin-1

Query Match 14.6%; Score 6; DB 2; Length 38;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 10 LFDPOH 13
 10 LFDPOH 15

RESULT 8
 conserved hypothetical protein PA4738 [Imported] - Pseudomonas aeruginosa (strain PA01)
 C/Species: Pseudomonas aeruginosa
 C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
 C/Accession: F83054
 R/Reviewer: C.K.; Pham, X.Q.; Ervin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,
 .; Loay, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A/Reference number: A82950; MUID:20437337; PMID:10984043
 A/Accession: F83054
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-65 <STO>
 A/Cross-references: UNIPROT:Q9HV61; GB:AE004887; GB:AE004091; NID:g9950991; PIDN:AA0812
 A/Experimental source: strain PA01
 C/Genetics: C/Genetic8
 A/Gene: PA4738

Query Match 14.6%; Score 6; DB 2; Length 65;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 16 SKERAE 21
 50 SKERAE 55

RESULT 9
 H+-transporting two-sector ATPase (EC 3.6.3.14) epsilon chain (atpc) - Bacillus subtilis
 C/Species: Bacillus subtilis
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
 C/Accession: I40369; G69591; G39257
 R/Santana, M.; Ionescu, M.S.; Vertes, A.; Longin, R.; Kuntz, F.; Danchin, A.; Glaser, P.
 J. Bacteriol. 176, 6802-6811, 1994
 A/Title: Bacillus subtilis FOF1 ATPase: DNA sequence of the atp operon and characterizat
 A/Reference number: I40360; MUID:95050246; PMID:7961438
 A/Accession: I40369
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-132 <RES>
 A/Cross-references: UNIPROT:P37812; EMBL:Z28592; NID:g433983; PIDN:CA82261.1; PID:g4339
 R/Kuntz, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
 C.; Bron, S.; Brouillet, S.; Brunsch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gallier
 lech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koeter, P.; Konigseisen, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinis
 A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
 Y.; M.; Ogawa, K.; Ogikawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanton,
 A/Authors: Schleif, S.; Schroeter, R.; Scottone, F.; Sekiguchi, J.; Sekowaka, A.; Serot
 akenchi, M.; Tamakoshi, A.; Tanaka, T.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A/Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A/Reference number: A69580; MUID:9604033; PMID:9386377
 A/Accession: G69591
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA

A/Residues: 1-132 <KUN>
 A/Cross-references: GB:Z99122; GB:AL009126; NID:g2636029; PIDN:CA815697.1; PID:g2636205
 A/Experimental source: strain 168
 C/Genetics:
 A/Gene: atpc
 C/Superfamily: H+-transporting ATP synthase epsilon chain
 C/Keywords: ATP biosynthesis; hydrolase; membrane-associated complex

Query Match 14.6%; Score 6; DB 1; Length 132;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 KERAE 22
 90 KERAE 95

RESULT 10
 D39384
 finger protein HRP6 - human (fragment)
 C/Species: Homo sapiens (man)
 C/Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 08-Dec-2000
 C/Accession: D39384
 R/Bellefroid, E.J.; Poncelet, D.A.; Lecocq, P.J.; Revelant, O.; Martial, J.A.
 Proc. Natl. Acad. Sci. U.S.A. 88, 3608-3612, 1991
 A/Title: The evolutionarily conserved Kruppel-associated box domain defines a subfamily
 A/Reference number: A39384; MUID:91219421; PMID:2023909
 A/Accession: D39384
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-154 <BEL>
 A/Cross-references: GB:M61869; NID:g184449; PID:g184450
 C/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
 C/Keywords: DNA binding; zinc finger

Query Match 14.6%; Score 6; DB 2; Length 154;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 5 KIFLFD 10
 102 KIFLFD 107

RESULT 11
 conserved hypothetical protein aq_987 - Aquifex aeolicus
 C/Species: Aquifex aeolicus
 C/Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
 C/Accession: E70385
 R/Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
 V.
 Nature 392, 353-358, 1998
 A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A/Reference number: A70300; MUID:98196666; PMID:9537320
 A/Accession: E70385
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-165 <AQF>
 A/Cross-references: UNIPROT:O67112; GB:AE000717; NID:g2983492; PIDN:AA07079.1; PID:g298
 A/Experimental source: strain VF5
 C/Genetics:
 A/Gene: aq_987
 C/Superfamily: Pyrococcus abyssi hypothetical protein PAB0445

Query Match 14.6%; Score 6; DB 2; Length 165;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 34 RLGSEE 39
 38 RLGSEE 43

RESULT 12

AE1538

transcription regulator Tetr/Acr family homolog 11n0845 [imported] - *Listeria innocua*C/Species: *Listeria innocua*

C/Date: 21-Nov-2001

C/Accession: AE1538

R/Glaeser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bioecker

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A/Author: Kieft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma

Ok, C.; Schluter, T.; Simoes, N.; Tixeront, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,

A/Title: Comparative genomics of *Listeria* species.

A/Reference number: AB1077; PMID:11679669

A/Accession: AE1538

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-184 <GLA>

A/Cross-references: UNIPROT:Q92D68; GB:AL592022; PIDN:CAC96077.1; PID:g16413296; GSPDB:C

A/Experimental source: strain C1jpl1262

C/Genetics:

A/Gene: 11n0845

Query Match

Best Local Similarity 14.6%; Score 6; DB 2; Length 184;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 35 LGEEK 40

73 LGEEK 78

RESULT 13

T1208

hypothetical protein F21D9.8 - *Caenorhabditis elegans*C/Species: *Caenorhabditis elegans*

C/Date: 15-Oct-1999

C/Accession: T1208

R/Basham, V.

submitted to the EMBL Data Library, November 1996

A/Reference number: Z19390

A/Accession: T1208

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-188 <W11>

A/Cross-references: UNIPROT:Q9XV69; EMBL:Z81510; PIDN:CAB04165.1; GSPDB:GN00023; CESP:F2

A/Experimental source: clone F21D9

C/Genetics:

A/Gene: CESP:F21D9.8

A/Map position: 5

A/Introns: 20/1; 90/3

Query Match

Best Local Similarity 14.6%; Score 6; DB 2; Length 188;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 18 EREAL 23

115 EREAL 120

RESULT 14

G65039

hypothetical protein b2612 - *Escherichia coli* (strain K-12)C/Species: *Escherichia coli*

C/Date: 12-Sep-1997

C/Accession: G65039

R/Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

Science 277, 1453-1462, 1997

A/Title: The complete genome sequence of *Escherichia coli* K-12.

A/Reference number: A64720; PMID:97426617; PMID:9278503

A/Accession: G65039

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-196 <BLAT>

A/Cross-references: GB:AE000347; GB:U00096; NID:G2367142; PIDN:AACT5661.1; PID:g1788965;

A/Experimental source: strain K-12, substrain MG1655

Query Match

Best Local Similarity 14.6%; Score 6; DB 2; Length 196;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 23 LMPLV 28

130 LMPLV 135

RESULT 15

F90919

repressor for uid operon [imported] - *Escherichia coli* (strain O157:H7, substrain RMD 0:C/Species: *Escherichia coli*

C/Date: 18-Jul-2001

C/Accession: F90919

R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genom

A/Reference number: A99629; PMID:1156231; PMID:11258796

A/Accession: F90919

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-196 <HAY>

A/Cross-references: UNIPROT:O59431; GB:BA000007; PIDN:BA835749.1; PID:g13361793; GSPDB:G

A/Experimental source: strain O157:H7, substrain RMD 050952

C/Genetics:

A/Gene: EC82326

Query Match

Best Local Similarity 14.6%; Score 6; DB 2; Length 196;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 13 HFISKE 18

51 HFISKE 56

Search completed: February 1, 2005, 15:33:06

Job time : 13.5556 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:54:06 ; Search time 70.5556 Seconds
(without alignments)
334.352 Million cell updates/sec

Title: SEQ5GLU2017
Perfect score: 41
Sequence: 1 NCLYKIFLPTQHFISKERALMPLVDQLENNRGSEKRF 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1825181 seqs, 575374646 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1810864

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Uniprot_02.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	958	1 BP28 MACFA	Q9GM44 macaca fasc
2	41	100.0	2144	1 BP28 HUMAN	Q9H583 homo sapien
3	41	100.0	2144	1 BP28 HUMAN	Q9H583 homo sapien
4	22	53.7	408	2 Q8VCK1	Q8VCK1 mus musculu
5	22	53.7	408	2 Q8VCK1	Q8VCK1 mus musculu
6	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
7	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
8	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
9	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
10	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
11	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
12	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
13	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
14	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
15	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
16	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
17	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
18	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
19	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
20	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
21	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
22	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
23	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
24	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
25	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
26	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
27	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
28	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
29	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
30	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu
31	20	48.8	897	2 Q8VCK1	Q8VCK1 mus musculu

32	7	17.1	339	2	Q7SD24	Q7SD24 neurospora
33	7	17.1	354	2	Q74975	Q74975 echizosacch
34	7	17.1	396	2	Q38033	Q38033 bacterioph
35	7	17.1	421	2	Q7MA16	Q7MA16 wolinnella s
36	7	17.1	422	2	Q8XNG2	Q8XNG2 clostridium
37	7	17.1	430	2	Q7ZTQ3	Q7ZTQ3 xenopus lae
38	7	17.1	457	2	Q8DKQ2	Q8DKQ2 synechococc
39	7	17.1	479	2	Q9M6E4	Q9M6E4 nicotiana t
40	7	17.1	479	2	Q8KYY0	Q8KYY0 uncultured
41	7	17.1	563	1	LTRB_LACIA	Q48722 lactococcus
42	7	17.1	563	1	LTRB_LACIA	Q48665 lactococcus
43	7	17.1	651	2	P93616	P93616 tritium ae
44	7	17.1	653	2	Q86924	Q86924 dictyosteli
45	7	17.1	658	2	Q9AT32	Q9AT32 daucus caro

ALIGNMENTS

```

RESULT 1
BP28_MACFA STANDARD; PRT; 958 AA.
ID BP28_MACFA
AC Q9GM44;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Protein BAP28 (Qnpa-17571) (Fragment).
GN Name-BAP28;
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the BAP28 family.
CC -1- SIMILARITY: Contains 1 HEAT repeat.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC
CC -----
CC EMBL; AB049842; BAB16728.1; ALT_INIT.
CC InterPro; IPR008938; ARM.
CC InterPro; IPR000357; HEAT.
CC PROSITE; PS50077; HEAT_REPEAT; FALSE_NEG.
CC
CC NOVELTER
CC REPEAT 920 956 HEAT.
CC SEQUENCE 958 AA; 108644 MW; 3DBD95C3623CFB31 CRC64;
Query Match 100.0%; Score 41; DB 1; Length 958;
Best Local Similarity 100.0%; Pred. No. 2,4e-34;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NCLYKIFLPTQHFISKERALMPLVDQLENNRGSEKRF 41
Db 811 NCLYKIFLPTQHFISKERALMPLVDQLENNRGSEKRF 851

```

RESULT 2

BP28_HUMAN STANDARD; PRT; 2144 AA.
AC Q9H5B3; Q9NM23;

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Protein BAP28.
 GN Name=BAP28;
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A., AND VARIANTS SER-1694; ALA-1854; ASP-1967 AND
 RP GLY-2017.
 RA Bougellet L., Chumakov I., Barry C., Cohen-Akenine A.;
 RT "A novel BAP28 gene and protein."
 RL Patent number WO0100669, 04-JAN-2001.
 [2]
 RP SEQUENCE OF 1534-2144 FROM N.A.
 RA Cobley V.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 [3]
 RN SEQUENCE OF 1777-2144 FROM N.A.
 RX PubMed=14702039; DOI=10.1038/ng1285;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Wakamatsu A., Hayashi K., Sato H., Negai K., Kimura K., Makita H.,
 RA Sekine M., Oobayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J., Ii, Saito K., Kawai Y., Isono Y., Nakamura Y.,
 RA Nagahari A., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
 RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
 RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
 RA Omura Y., Abe K., Kamihara K., Katsura N., Sato K., Tanikawa M.,
 RA Yamazaki M., Ninomiya K., Ishikashi T., Yamashita H., Murakawa K.,
 RA Fujimori K., Tanai H., Kimura M., Watanabe M., Hirotsu K., Chiba Y.,
 RA Ishida S., Oono Y., Takiguchi S., Watanabe S., Yosida M., Huchita T.,
 RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
 RA Nomura Y., Togaiya S., Komai F., Hara R., Takeuchi K., Arita M.,
 RA Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotaka S.,
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
 RA Moriya S., Montiyama H., Satoh N., Takami S., Terahashi Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 RA Hishigaki H., Watanabe K., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
 RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Maeno Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
 RT "Complete sequencing and characterization of 21,243 full-length human
 cDNAs."
 RL Nat. Genet. 36:40-45 (2004).
 CC -1 SIMILARITY: Belongs to the BAP28 family.
 CC -1 SIMILARITY: Contains 1 HEAT repeat.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).

FT

KW REPEAT 2106 2142

HEAT.

FT VARIANT 1694 1694 N -> S.
 FT FT /FTid=VAR_010939.
 FT VARIANT 1854 1854 V -> A.
 FT FT /FTid=VAR_010940.
 FT VARIANT 1967 1967 N -> D.
 FT FT /FTid=VAR_010941.
 FT VARIANT 2017 2017 E -> G.
 FT FT /FTid=VAR_010942.
 SQ SEQUENCE 2144 AA, 242355 MW, D66816EE78D8C9B7 CRC64;
 Query Match 100.0%; Score 41; DB 1; Length 2144;
 Best Local Similarity 100.0%; Pred. No. 4,5e-34;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 NCXYKIFLFDYQHFISKEAEALMPLVDQLENRLGSEKRF 41
 Db 1997 NCXYKIFLFDYQHFISKEAEALMPLVDQLENRLGSEKRF 2037
 RESULT 3
 ID 08VCX1 PRELIMINARY; PRT; 349 AA.
 AC 08VCX1;
 DT 01-MAR-2002 (TEMBLrel. 20, Created)
 DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
 DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
 DE BC019693 protein.
 GN Name=BC019693;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=mix FVB/N;
 RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strasserberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Sheman C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buecaw K.H., Scheffer C.F., Bat N.K.,
 RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Stachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalski U., Smallus D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Merra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=mix FVB/N;
 RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
 RA Strussberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC019693; AAH19693.1; -
 DR MGD; MGI:2384983; BC019693.
 DR InterPro; IPR008938; ARM.
 SQ SEQUENCE 349 AA, 40166 MW, 9763D0331AD0F515 CRC64;
 Query Match 53.7%; Score 22; DB 2; Length 349;
 Best Local Similarity 100.0%; Pred. No. 1.1e-14;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 18 ERAEALMPLVDQLENRLGSE 39

Db 219 ERAEALMPLVDQLENRLGSEE 240

RESULT 4

Q8CCT5 PRELIMINARY; PRT; 408 AA.
 ID Q8CCT5
 AC Q8CCT5
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Mus musculus adult male olfactory brain cDNA, RIKEN full-length
 DE enriched library, clone:643040D06 product:hypothetical AFM repeat
 DE structure containing protein, full insert sequence.
 GN Neme-B019693;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
 RX MEDLINE=99279253; PubMed=10349636;
 RA Carninci P., Hayashizaki Y.,
 RT "High-efficiency full-length cDNA cloning.",
 RL Meth. Enzymol. 303:19-44(1999).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
 RX MEDLINE=21085660; PubMed=11217851;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.",
 RL Nature 409:685-690(2001).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
 RA The FANTOM Consortium,
 RT "The RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.",
 RL Nature 420:563-573(2002).
 [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
 RX MEDLINE=20499374; PubMed=11042153;
 RA Kanno H., Okazaki Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RT "Normalization and subtractions of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.",
 RL Genome Res. 10:1617-1630(2000).
 [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
 RX MEDLINE=20530913; PubMed=11076861;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Kanno H., Akiyama J., Nishi K., Kiteuna T., Tashiro H., Itoh M.,
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto K., Matsuno H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujimake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
 RT "RIKEN integrated sequence analysis (RISA) system-364-format
 RT sequencing pipeline with 364 multicapillary sequencer.",
 RL Genome Res. 10:1757-1771(2000).
 [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
 RA Adachi J., Aizawa K., Akimura T., Arikawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hirooka K., Hirozane T.,
 RA Horii F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Kondo H., Koyama M., Koya S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,

RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohnato N., Okazaki Y.,
 RA Saito R., Saitoh H., Sakai C., Sakai C., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akaiira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK032134; BAC27721.1;
 DR MGD; MGI:2384983; BC019693.
 DR InterPro; IPR008938; ARM.
 KN Hypothetical protein.
 SQ SEQUENCE 408 AA; 4668 MW; 42AB9EB13CA3F6E7 CRC64;

Query Match 53.7%; Score 22; DB 2; Length 408;
 Best Local Similarity 100.0%; Pred. No. 1.2e-14;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 ERAEALMPLVDQLENRLGSEE 39
 Db 278 ERAEALMPLVDQLENRLGSEE 299

RESULT 5

Q96ES5 PRELIMINARY; PRT; 349 AA.
 ID Q96ES5
 AC Q96ES5
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE FLJ10359 protein.
 GN Name=FLJ10359;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Hsieh F.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udén T.B., Toehiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huylk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Holton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butcherfield Y.S.,
 RA Krzywnicki M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.",
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; BC011983; AAH11983.1;
 DR InterPro; IPR008938; ARM.
 SQ SEQUENCE 349 AA; 39921 MW; 3A359597FF7079EB CRC64;

Query Match 48.8%; Score 20; DB 2; Length 349;
 Best Local Similarity 100.0%; Pred. No. 1.4e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NCYKIFLPTDHFISKERA 20
 |||||

Db 202 NCLYKIFLPTDHFISKERA 221

RESULT 6

ID Q8N7L7 PRELIMINARY; PRT; 897 AA.

AC Q8N7L7; 01-OCT-2002 (TREMBlrel. 22, Created)

DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Hypothetical protein FLJ40893.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

[1]

SEQUENCE FROM N.A.

RC TISSUE=Lymph;

RX PubMed=14702039;

RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,

RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,

RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,

RA Yamamoto K., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,

RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,

RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,

RA Takekoshi K., Ishihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,

RA Niimiya K., Ishihara T., Yamashita H., Murakawa K., Fujimori K.,

RA Tanai H., Kimata M., Matsumoto S., Hirao S., Chiba Y., Iehida S.,

RA Ono Y., Takiguchi S., Matsumoto S., Yoshida M., Hattori T., Kusano J.,

RA Kanehori K., Takahashi F., Hara R., Takeuchi K., Arita M., Imose N.,

RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,

RA Mutsaers K., Yuki H., Oshima A., Sasaki N., Aotaka S.,

RA Moriya S., Momiya H., Sato N., Takami S., Terashima Y., Suzuki O.,

RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shindzu F., Makabe H.,

RA Hishigaki H., Matsumoto T., Sugiyama A., Takemoto M., Kawakami B.,

RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,

RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,

RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,

RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senda T.,

RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,

RA Mizushima-Sugano J., Sato T., Shirai Y., Takahashi Y., Nakagawa K.,

RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,

RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,

RT "Complete sequencing and characterization of 21,243 full-length human

RT cDNAs."

RL Nat. Genet. 36:40-45 (2004).

DR EMBL, AK098212; BAC05261.1; -

DR InterPro, IPR008938; ARM.

SO SEQUENCE 897 AA; 101574 MW; 5FF6A94FE8855895 CRC64;

Query Match 48.8%; Score 20; DB 2; Length 897;

Best Local Similarity 100.0%; Pred. No. 2,9e-12;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NCLYKIFLPTDHFISKERA 20

Db 750 NCLYKIFLPTDHFISKERA 769

RESULT 7

ID Q6P197 PRELIMINARY; PRT; 1106 AA.

AC Q6P197; 05-JUL-2004 (TREMBlrel. 27, Created)

DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)

DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)

DE FLJ10359 protein (Fragment).

OS Homo sapiens (Human).

GN Name=FLJ10359;

RA Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

[1]

SEQUENCE FROM N.A.

RC TISSUE=Lymph;

RX PubMed=12477932;

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marisano K., Farmer A.A., Rubin G.M., Hong L.,

RA Stadelton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schaefer T.E.,

RA Brownstein M.J., Udén T.B., Toehyuk S., Carninci P., Prange C.,

RA Rana S.S., Loggiano N.A., Peters G.J., Abramson R.D., Mullany S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton A., Kettman A.C., Shevchenko Y., Bouffard G.G.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Kravynski M.I., Skalek U., Smallus D.E., Schnerch A., Schein J.E.,

RA Jones S.J., Marra M.A.,

RT "Generation and initial analysis of more than 15,000 full-length human

RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

[2]

SEQUENCE FROM N.A.

RC TISSUE=Lymph;

RA Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.

DR EMBL, BC065205; AAH65205.1; -

DR InterPro, IPR008938; ARM.

SO NON TER 1 125359 MW; 09F7CE94042302C4 CRC64;

Qy 1 NCLYKIFLPTDHFISKERA 20

Db 959 NCLYKIFLPTDHFISKERA 978

RESULT 8

ID AAH65205 PRELIMINARY; PRT; 1106 AA.

AC AAH65205; 02-MAR-2004 (TREMBlrel. 27, Created)

DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)

DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)

DE FLJ10359 protein (Fragment).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

[1]

SEQUENCE FROM N.A.

RC TISSUE=Lymph;

RX PubMed=12477932;

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marisano K., Farmer A.A., Rubin G.M., Hong L.,

RA Stadelton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schaefer T.E.,

RA Brownstein M.J., Udén T.B., Toehyuk S., Carninci P., Prange C.,

RA Rana S.S., Loggiano N.A., Peters G.J., Abramson R.D., Mullany S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,

```

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Keeteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.N., Butterfield Y.S.,
RA Krzywinski M.I., Skalske U., Smalins D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Maira M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN
RN [2]
RN SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC065205; AAH65205.1; -.
FT NON TER 1
SQ SEQUENCE 1106 AA; 125359 MW; 09F7CE94042302C4 CRC64;

Query Match 48.8%; Score 20; DB 2; Length 1106;
Best Local Similarity 100.0%; Pred. No. 3,4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NCLYKIFLDTQHFISKERA 20
Db 959 NCLYKIFLDTQHFISKERA 978

RESULT 9
Q72BK8 PRELIMINARY; PRT; 241 AA.
ID Q72BK8
DT 05-JUN-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUN-2004 (TReMBLrel. 27, Last annotation update)
DE ABC transporter, ATP-binding protein.
GN OrderedLocusNames=DVU1627;
OS Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB
OS 8303).
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
OC Desulfovibrionaceae; Desulfovibrio.
OX NCBI_TaxID=882;
RN
RN [1]
RN SEQUENCE FROM N.A.
RX PubMed=15077118; DOI=10.1038/nbt959;
RA Heidelberg J.F., Seehardt R., Haveman S.A., Hemme C.L., Paulsen I.T.,
RA Kolonay J.F., Eisen J.A., Ward N.L., Melhe B.A., Brinkac L.M.,
RA Daugherty S.C., Debey R.T., Dodson R.J., Durkin A.S., Madupu R.,
RA Nelson W.C., Sullivan S.A., Fouts D.E., Haft D.H., Selengut J.,
RA Peterson J.D., Davidson T.M., Zafar N., Zhou L., Radune D.,
RA Dmitrov G., Hance M., Tran K., Khouri H.M., Gill J., Uitterback T.R.,
RA Feldblyum T.V., Wall J.D., Voordouw G., Fraser C.M.;
RT "The genome sequence of the anaerobic, sulfate-reducing bacterium
RT Desulfovibrio vulgaris Hildenborough.";
RL Nat. Biotechnol. 22:554-559 (2004).
CC -1- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL; AE017314; AAS96105.1; -.
DR TIGR; DVU1627; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR InterPro; IPR003593; AAA_Arpase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA_1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KM ATP-binding; Complete proteome.
SQ SEQUENCE 241 AA; 26639 MW; B9A37B368194F6FD CRC64;

Query Match 19.5%; Score 8; DB 2; Length 241;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 17 KERAALM 24
Db 115 KERAALM 122

RESULT 10
AAS96105 PRELIMINARY; PRT; 241 AA.
ID AAS96105
AC AAS96105;
DT 27-APR-2004 (TReMBLrel. 27, Created)
DT 27-APR-2004 (TReMBLrel. 27, Last sequence update)
DT 11-MAY-2004 (TReMBLrel. 27, Last annotation update)
DE ABC transporter, ATP-binding protein.
GN DVU1627.
OS Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB
OS 8303).
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
OC Desulfovibrionaceae; Desulfovibrio.
OX NCBI_TaxID=882;
RN
RN [1]
RN SEQUENCE FROM N.A.
RX PubMed=15077118;
RA Heidelberg J.F., Seehardt R., Haveman S.A., Hemme C.L., Paulsen I.T.,
RA Kolonay J.F., Eisen J.A., Ward N., Melhe B.A., Brinkac L.M.,
RA Daugherty S.C., Debey R.T., Dodson R.J., Durkin A.S., Madupu R.,
RA Nelson W.C., Sullivan S.A., Fouts D.E., Haft D.H., Selengut J.,
RA Peterson J.D., Davidson T.M., Zafar N., Zhou L., Radune D.,
RA Dmitrov G., Hance M., Tran K., Khouri H.M., Gill J., Uitterback T.R.,
RA Feldblyum T.V., Wall J.D., Voordouw G., Fraser C.M.;
RT "The genome sequence of the anaerobic, sulfate-reducing bacterium
RT Desulfovibrio vulgaris Hildenborough.";
RL Nat. Biotechnol. 22:554-559 (2004).
DR EMBL; AE017314; AAS96105.1; -.
DR TIGR; DVU1627; -.
DR KM ATP-binding.
SQ SEQUENCE 241 AA; 26639 MW; B9A37B368194F6FD CRC64;

Query Match 19.5%; Score 8; DB 2; Length 241;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 KERAALM 24
Db 115 KERAALM 122

RESULT 11
Q7T153 PRELIMINARY; PRT; 1278 AA.
ID Q7T153
AC Q7T153;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE SI:ZC146F4.2.2 (Novel protein similar to human BAP28) (Fragment).
GN Name=SI:ZC146F4.2;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN
RN [1]
RN SEQUENCE FROM N.A.
RA Garner P.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL732629; CAB7602.1; -.
FT NON TER 1
SQ SEQUENCE 1278 AA; 143523 MW; E6C9FC81B77EE1A9 CRC64;

Query Match 19.5%; Score 8; DB 2; Length 1278;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

ID	Accession	Length	Score	DB 2	Length	Score	DB 2
DB	1214 PLVNDLEN 1221	1221	19.5%	Score 8;	DB 2;	Length 1336;	1336
QY	26 PLVNDLEN 33	33	100.0%	Pred. No. 17;			17
DB	1214 PLVNDLEN 1221	1221	0	Mismatches	0	Indels	0
QY	26 PLVNDLEN 33	33	0	Gaps	0		0
RESULT 13							
Q2SY48	PRELIMINARY;	PRT;	2159	AA.			
ID	Q7SY48						
AC	Q7SY48;						
DT	01-OCT-2003 (TREMBlrel. 25, Created)						
DT	01-OCT-2003 (TREMBlrel. 25, Last sequence update)						
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)						
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)						
DE	SI:2C146F4.2.1 (Novel protein similar to human BAP28) (Fragment).						
OS	Brachydanio rerio (zebrafish) (Danio rerio).						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;						
OC	Cyprinidae; Danio.						
OK	NCB1_TaxID=7955;						
RN	(1)						
RP	SEQUENCE FROM N.A.						
RA	Garner P.;						
RL	Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.						
DR	EMBL; AL732623; CAC17603.1; ..						
FT	NON_TER	1					
SO	SEQUENCE 1336 AA; 150326 MW; AA949557F21ACBCE CRC64;						
Query Match			19.5%	Score 8;	DB 2;	Length 1336;	1336
Best Local Similarity			100.0%	Pred. No. 17;			17
Matches	8;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						0;

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=whole body;
RA Strausberg R.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
RW EMBL; BC055128; AAH55128.1; -.
KW Hypothetical protein.
SQ SEQUENCE 2159 AA; 242048 MW; F00DBBNAJD597E70B CRC64;

Query Match 19.5%; Score 8; DB 2; Length 2159;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 LVDQLEN 33
DB 2037 LVDQLEN 2044

RESULT 14
EX75_VIBPA
ID EX75_VIBPA STANDARD; PRT; 80 AA.
AC Q87R9;
DT 10-OCT-2003 (Rel. 42, Created)
DI 10-OCT-2003 (Rel. 42, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable exodeoxyribonuclease VII small subunit (EC 3.1.11.6)
DN (Name=xseB; OrderedlocusNames=VP0688;
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIMD 221063 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620729;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Najima W., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yaenagata T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae.";
RL Lancet 361:743-749(2003).
CC -1- FUNCTION: Bidirectionally degrades single-stranded DNA into large
CC acid-insoluble oligonucleotides, which are then degraded further
CC into small acid-soluble oligonucleotides (By similarity).
CC -1- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5'- to 3'-
CC or 3'- to 5'-direction to yield nucleoside 5'-phosphates.
CC -1- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the xseB family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP005075; BAC8951.1; -.
CC DR HAMAP; MF 00337; -. 1.
CC DR InterPro; IPR003761; Exonuc_VII_S.
CC pfam; PF02609; Exonuc_VII_S.1.
CC DR Complete proteome; Exonuclease; Hydroxylase; Nuclease.
CC KW SEQUENCE 80 AA; 8879 MW; 03B881D6B8EC4F6A CRC64;
QY
Query Match 17.1%; Score 7; DB 1; Length 80;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 27 LVDQLEN 33
|||||

```

Db 21 LVDQLEN 27

RESULT 15

EX7S_VIBVU STANDARD; PRT; 80 AA.

AC Q8DFR5;

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Probable exodeoxyribonuclease VII small subunit (EC 3.1.11.6)

DE (Exonuclease VII small subunit).

GN Name=xseB; OrderedLocName=VV10313;

OS Vibrio vulnificus.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

OC Vibrionaceae; Vibrio.

OX NCBI_TaxID=672;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CMCP6;

RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,

RA Choy H.E.;

RT "Complete genome sequence of Vibrio vulnificus CMCP6.";

RL Submitted (DEC-2002) to the EMBL/Genbank/DBJ databases.

CC -1 FUNCTION: Bidirectionally degrades single-stranded DNA into large

acid-insoluble oligonucleotides, which are then degraded further

into small acid-soluble oligonucleotides (by similarity).

CC -1 CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5' to 3' or 3' to 5' direction to yield nucleoside 5'-phosphates.

CC -1 SUBUNIT: Heterooligomer composed of large and small subunits (by similarity).

CC -1 SUBCELLULAR LOCATION: Cytoplasmic (by similarity).

CC -1 SIMILARITY: Belongs to the xseB family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AE016798; AAC08843.1; -.

DR HAMAP; MF_00337; -; 1.

DR InterPro; IPR003761; Exonuc_VII_S.

DR Pfam; PF02609; Exonuc_VII_S; 1.

DR TIGRFAMs; TIGR01280; xseB; 1.

KW Complete proteome; Exonuclease; Hydrolase; Nuclease.

SO SEQUENCE 80 AA; 8756 MW; FE23B015BD49F5ED CRC64;

Query Match 17.1%; Score 7; DB 1; Length 80;

Best Local Similarity 100.0%; Pred. No. 21;

Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 27 LVDQLEN 33

Db 21 LVDQLEN 27

Search completed: February 1, 2005, 15:31:07
Job time : 71.5556 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model1

Run on: February 1, 2005, 14:58:28 / Search time 17.5556 Seconds
(Without alignments)
154.882 Million cell updates/sec

Title: SEQ5GLY2017

Perfect score: 41

Sequence: 1 NCILYKIFLPTDQHFISKERA.....ALMPLVDQLENLGEKRF 41

Scoring table: OLIGO

Searched: 478139 seqs, 66318000 residues

Word size: 0

Total number of hits satisfying chosen parameters: 290433

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/PTUS_COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	17.1	428	3 US-09-347-833-6	Sequence 6, Appli
2	7	17.1	650	4 US-09-489-039A-9271	Sequence 9271, Ap
3	7	17.1	655	3 US-09-347-833-4	Sequence 4, Appli
4	6	14.6	69	4 US-09-621-976-6451	Sequence 6451, Ap
5	6	14.6	80	4 US-09-248-796A-22366	Sequence 22366, A
6	6	14.6	95	4 US-09-513-999C-5651	Sequence 5651, Ap
7	6	14.6	107	3 US-09-087-465-22	Sequence 22, Appli
8	6	14.6	141	3 US-09-526-542-4	Sequence 4, Appli
9	6	14.6	141	3 US-09-526-542-6	Sequence 4, Appli
10	6	14.6	141	4 US-10-117-087-6	Sequence 6, Appli
11	6	14.6	141	4 US-10-117-087-6	Sequence 6, Appli
12	6	14.6	161	5 PCT-US95-05741-11	Sequence 11, Appli
13	6	14.6	181	2 US-08-482-142-193	Sequence 193, App
14	6	14.6	181	2 US-08-478-572-193	Sequence 193, App
15	6	14.6	181	3 US-08-484-296-193	Sequence 193, App
16	6	14.6	195	2 US-08-883-704A-2	Sequence 2, Appli
17	6	14.6	195	2 US-09-151-957-2	Sequence 2, Appli
18	6	14.6	195	4 US-10-195-158-2	Sequence 2, Appli
19	6	14.6	227	1 US-08-254-493-1	Sequence 1, Appli
20	6	14.6	227	2 US-08-253-751-6	Sequence 6, Appli
21	6	14.6	227	2 US-08-453-925-6	Sequence 6, Appli
22	6	14.6	227	3 US-08-403-253A-6	Sequence 6, Appli
23	6	14.6	227	4 US-08-435-816A-6	Sequence 6, Appli
24	6	14.6	228	1 US-08-408-222B-1	Sequence 1, Appli
25	6	14.6	235	4 US-09-270-767-57304	Sequence 57304, A
26	6	14.6	236	4 US-09-107-532A-6816	Sequence 6816, Ap
27	6	14.6	236	4 US-09-134-000C-5104	Sequence 5104, Ap

28	6	14.6	242	4 US-09-252-991A-17571	Sequence 17571, A
29	6	14.6	259	4 US-09-252-991A-26709	Sequence 26709, A
30	6	14.6	274	4 US-09-248-796A-24969	Sequence 24969, A
31	6	14.6	281	4 US-09-270-767-58539	Sequence 58539, A
32	6	14.6	290	4 US-09-489-039A-11474	Sequence 11474, A
33	6	14.6	293	4 US-09-270-767-43443	Sequence 43443, A
34	6	14.6	305	4 US-09-328-352-7116	Sequence 7116, Ap
35	6	14.6	306	4 US-09-252-991A-25195	Sequence 25195, A
36	6	14.6	333	4 US-09-543-681A-7983	Sequence 7983, Ap
37	6	14.6	352	2 US-08-483-926A-11	Sequence 11, Appli
38	6	14.6	352	2 US-08-737-045-12	Sequence 12, Appli
39	6	14.6	368	4 US-09-252-991A-17027	Sequence 17027, A
40	6	14.6	368	4 US-09-248-796A-14941	Sequence 14941, A
41	6	14.6	376	3 US-09-387-418A-13	Sequence 13, Appli
42	6	14.6	383	4 US-09-252-991A-25307	Sequence 25307, A
43	6	14.6	387	4 US-09-489-039A-14027	Sequence 14027, A
44	6	14.6	393	3 US-09-387-418A-10	Sequence 10, Appli
45	6	14.6	398	4 US-09-252-991A-19301	Sequence 19301, A

ALIGNMENTS

RESULT 1
US-09-347-833-6
Sequence 6, Application US/09347833

Patent No. 6294658

GENERAL INFORMATION:

APPLICANT: Famodu, Layo O.

TITLE OF INVENTION: Factors Involved in Gene Expression

FILE REFERENCE: BB-1172

CURRENT FILING DATE: 1999-07-02

EARLIER APPLICATION NUMBER: 60/092,415

NUMBER OF SEQ ID NOS: 11

SOFTWARE: Microsoft Office 97

SEQ ID NO 6

LENGTH: 428

TYPE: PRT

ORGANISM: Glycine max

US-09-347-833-6

Query Match 17.1%; Score 7; DB 3; Length 428;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 PLYDQLE 32

357 PLYDQLE 363

RESULT 2
US-09-489-039A-9271
Sequence 9271, Application US/09489039A

Patent No. 6610836

GENERAL INFORMATION:

APPLICANT: Gary Breton et. al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

FILE REFERENCE: 2709.2004001

CURRENT FILING DATE: 2000-01-27

PRIOR APPLICATION NUMBER: US 60/117,747

NUMBER OF SEQ ID NOS: 14342

SEQ ID NO 9271

LENGTH: 650

TYPE: PRT

ORGANISM: Klebsiella pneumoniae

US-09-489-039A-9271

Query Match 17.1%; Score 7; DB 4; Length 650;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 GALLMPL 27
Db 9 GALLMPL 15

RESULT 3

US-09-347-833-4
; Sequence 4, Application US/09347833
; Patent No. 6294658
; GENERAL INFORMATION:
; APPLICANT: Famodu, Layo O.
; APPLICANT: Odeh, Joan T.
; TITLE OF INVENTION: Factors Involved in Gene Expression
; FILE REFERENCE: BB-1172
; CURRENT APPLICATION NUMBER: US/09/347,833
; CURRENT FILING DATE: 1999-07-02
; EARLIER APPLICATION NUMBER: 60/092,415
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-347-833-4

Query Match 17.1%; Score 7; DB 3; Length 655;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 PLVDQLE 32
Db 583 PLVDQLE 589

RESULT 4

US-09-621-976-6451
; Sequence 6451, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 6451
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-6451

Query Match 14.6%; Score 6; DB 4; Length 69;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LYKIFL 8
Db 57 LYKIFL 62

RESULT 5

US-09-248-796A-22566
; Sequence 22566, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANT
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 22566
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-22566

Query Match 14.6%; Score 6; DB 4; Length 80;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LYKIFL 8
Db 56 LYKIFL 61

RESULT 6

US-09-513-999C-5651
; Sequence 5651, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 5651
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-513-999C-5651

Query Match 14.6%; Score 6; DB 4; Length 95;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 AGALMM 25
Db 66 AGALMM 71

RESULT 7

US-09-087-465-22
; Sequence 22, Application US/09087465A
; Patent No. 6160092
; GENERAL INFORMATION:
; APPLICANT: Vinkemeier, Dwe
; APPLICANT: Chen, Xiaomln
; APPLICANT: Darnell Jr., James E
; APPLICANT: Kurlyan, John
; TITLE OF INVENTION: A CRYSTAL OF THE CORE PORTION OF A STAF AND METHODS OF
; FILE REFERENCE: 600-1-229
; CURRENT APPLICATION NUMBER: US/09/087,465A
; CURRENT FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 37

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 107
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-087-465-22
```

```
Query Match          14.6%; Score 6; DB 3; Length 107;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY          14 FISHER 19
           |||||
Db          5 FISHER 10
```

```
RESULT 8
US-09-526-542-4
; Sequence 4, Application US/09526542
; Patent No. 6369198
; GENERAL INFORMATION:
; APPLICANT: SERIUPI-CRESCENZI, Octaviano
; APPLICANT: DELLA PIETRA, Linda
; TITLE OF INVENTION: ALLELIC VARIANTS OF HUMAN STAT3
; FILE REFERENCE: SERIUPI-2
; CURRENT APPLICATION NUMBER: US/09/526,542
; CURRENT FILING DATE: 2000-03-19
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2)..(424)
; OTHER INFORMATION: note: "SH2 domain of the published hSTAT3 sequence (Akira et al.)"
US-09-526-542-4
```

```
Query Match          14.6%; Score 6; DB 3; Length 141;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY          14 FISHER 19
           |||||
Db          25 FISHER 30
```

```
RESULT 9
US-09-526-542-6
; Sequence 6, Application US/09526542
; Patent No. 6369198
; GENERAL INFORMATION:
; APPLICANT: SERIUPI-CRESCENZI, Octaviano
; APPLICANT: DELLA PIETRA, Linda
; TITLE OF INVENTION: ALLELIC VARIANTS OF HUMAN STAT3
; FILE REFERENCE: SERIUPI-2
; CURRENT APPLICATION NUMBER: US/09/526,542
; CURRENT FILING DATE: 2000-03-19
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Mouse
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2)..(424)
; OTHER INFORMATION: note: "SH2 domain of murine STAT3"
US-09-526-542-6
```

```
Query Match          14.6%; Score 6; DB 3; Length 141;
Best Local Similarity 100.0%; Pred. No. 48;
```

```
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY          14 FISHER 19
           |||||
Db          25 FISHER 30
```

```
RESULT 10
US-10-117-087-4
; Sequence 4, Application US/10117087
; Patent No. 6660848
; GENERAL INFORMATION:
; APPLICANT: SERIUPI-CRESCENZI, Octaviano
; APPLICANT: DELLA PIETRA, Linda
; TITLE OF INVENTION: ALLELIC VARIANTS OF HUMAN STAT3
; FILE REFERENCE: SERIUPI-2
; CURRENT APPLICATION NUMBER: US/10/117,087
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: US/09/526,542
; PRIOR FILING DATE: 2000-03-19
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2)..(424)
; OTHER INFORMATION: note: "SH2 domain of the published hSTAT3 sequence (Akira et al.)"
US-10-117-087-4
```

```
Query Match          14.6%; Score 6; DB 4; Length 141;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY          14 FISHER 19
           |||||
Db          25 FISHER 30
```

```
RESULT 11
US-10-117-087-6
; Sequence 6, Application US/10117087
; Patent No. 6660848
; GENERAL INFORMATION:
; APPLICANT: SERIUPI-CRESCENZI, Octaviano
; APPLICANT: DELLA PIETRA, Linda
; TITLE OF INVENTION: ALLELIC VARIANTS OF HUMAN STAT3
; FILE REFERENCE: SERIUPI-2
; CURRENT APPLICATION NUMBER: US/10/117,087
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: US/09/526,542
; PRIOR FILING DATE: 2000-03-19
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Mouse
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2)..(424)
; OTHER INFORMATION: note: "SH2 domain of murine STAT3"
US-10-117-087-6
```

```
Query Match          14.6%; Score 6; DB 4; Length 141;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY          14 FISHER 19
           |||||
```

DB 25 FISKER 30

RESULT 12

PCT-US95-05741-11
Sequence 11, Application PC/TUS9505741

GENERAL INFORMATION:

APPLICANT: Weintraub, Harold
APPLICANT: Lee, Jacqueline E.
APPLICANT: Tapscott, Stephen J.
APPLICANT: Hollenberg, Stanley M.
TITLE OF INVENTION: Neurogenic Differentiation (Neurod) Gene
TITLE OF INVENTION: and Protein
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen O'Connor Johnson Kindness
STREET: 1420 Fifth Avenue, Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05741
FILING DATE:
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Broderick, Thomas F.
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: FPCR-1-8504
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-8100
TELEFAX: 206-225-0709
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-05741-11

Query Match

Best Local Similarity 14.6%; Score 6; DB 5; Length 161;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;OY 34 RLGEE 39
|||

DB 33 RLGEE 38

RESULT 13

US-08-482-142-193
Sequence 193, Application US/08482142

Patent No. 5920862

GENERAL INFORMATION:

APPLICANT: Garman, Richard
APPLICANT: Greenstein, Julia
APPLICANT: Kuo, Mei-chang
APPLICANT: Rogers, Bruce
APPLICANT: Franzen, Henry
APPLICANT: Chen, Xian
APPLICANT: Evans, Sean
APPLICANT: Shaked, Ze'ev
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
STREET: 610 LINCOLN STREET

CITY: WALTHAM

STATE: MA
COUNTRY: USA
ZIP: 02154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,142
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/445,307
FILING DATE: 07 June 1995ATTORNEY/AGENT INFORMATION:
NAME: CRAIG, ANNE I.
REGISTRATION NUMBER: 32,976
REFERENCE/DOCKET NUMBER: 017,6US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 466-6000
TELEFAX: (617) 466-6040INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-482-142-193

Query Match

Best Local Similarity 14.6%; Score 6; DB 2; Length 181;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;OY 27 LVDOLE 32
|||

DB 72 LVDOLE 77

RESULT 14

US-08-478-572-193
Sequence 193, Application US/08478572

Patent No. 5968526

GENERAL INFORMATION:

APPLICANT: Garman, Richard
APPLICANT: Greenstein, Julia
APPLICANT: Kuo, Mei-chang
APPLICANT: Rogers, Bruce
APPLICANT: Franzen, Henry
APPLICANT: Chen, Xian
APPLICANT: Evans, Sean
APPLICANT: Shaked, Ze'ev
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
STREET: 610 LINCOLN STREET
CITY: WALTHAM
STATE: MA
COUNTRY: USA
ZIP: 02154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,572
FILING DATE: 07-June-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/445,307
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: CRAIG, ANNE I.
REGISTRATION NUMBER: 32,976
REFERENCE/DOCKET NUMBER: 017.6US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 466-6000
TELEFAX: (617) 466-6040
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 181 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-478-572-193

Query Match 14.6%; Score 6; DB 2; Length 181;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 27 LVDPLE 32
|||
Db 72 LVDPLE 77

RESULT 15
US-08-484-296-193
Sequence 193, Application US/08484296
Patent No. 6268491
GENERAL INFORMATION:
APPLICANT: Garman, Richard
APPLICANT: Greenstein, Julia
APPLICANT: Kuo, Mei-Chang
APPLICANT: Rogers, Bruce
APPLICANT: Franzen, Henry
APPLICANT: Chen, Xian
APPLICANT: Evans, Sean
APPLICANT: Shaked, Ze'ev
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
STREET: 610 LINCOLN STREET
CITY: WALTHAM
STATE: MA
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,296
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/445,307
FILING DATE: 07 June 1995
ATTORNEY/AGENT INFORMATION:
NAME: CRAIG, ANNE I.
REGISTRATION NUMBER: 32,976
REFERENCE/DOCKET NUMBER: 017.6US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 466-6000
TELEFAX: (617) 466-6040
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 181 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-484-296-193
Query Match 14.6%; Score 6; DB 3; Length 181;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 27 LVDPLE 32
|||
Db 72 LVDPLE 77

Search completed: February 1, 2005, 15:35:45
job time : 18.5556 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 15:09:39 / Search time 59.7778 Seconds
(without alignments)
247.799 Million cell updates/sec

Title: SEQ5GLY2017

Perfect score: 41

Sequence: 1 NCLYKIFLPDTQHFIKSKRA.....ALMPLVDQLENRLGSEKFK 41

Scoring table: OLIGO

Gapop 60.0, Gapext 60.0

Searched: 1608061 seqs, 361289366 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1311380

Minimum DB seq length: 21

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

Published Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*

3: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*

4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*

5: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*

6: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*

7: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*

8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*

9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*

10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*

11: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*

12: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*

13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*

14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*

15: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*

16: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*

17: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*

18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*

19: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*

20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	48.8	1149	17	US-10-128-558-167
2	7	17.1	53	15	US-10-424-599-275868
3	7	17.1	238	16	US-10-767-701-41775
4	7	17.1	314	16	US-10-437-963-187132
5	7	17.1	349	15	US-10-424-599-149524
6	7	17.1	406	15	US-10-425-114-53502
7	7	17.1	633	16	US-10-437-963-110339
8	7	17.1	660	16	US-10-437-963-151847
9	7	17.1	662	16	US-10-437-963-110342
10	7	17.1	663	15	US-10-424-599-149522
11	6	14.6	36	9	US-09-864-761-44493
12	6	14.6	45	17	US-10-425-115-340619
13	6	14.6	48	9	US-09-989-920-197

14	6	14.6	51	15	US-10-424-599-219715	Sequence 219715,
15	6	14.6	52	14	US-10-029-386-29247	Sequence 29247, A
16	6	14.6	55	17	US-10-425-115-203297	Sequence 203297,
17	6	14.6	61	15	US-10-424-599-245649	Sequence 245649,
18	6	14.6	71	15	US-10-282-122A-68905	Sequence 68905, A
19	6	14.6	72	16	US-10-437-963-178932	Sequence 178932,
20	6	14.6	73	17	US-10-425-115-266081	Sequence 266081,
21	6	14.6	74	17	US-10-425-115-364110	Sequence 364110,
22	6	14.6	83	9	US-09-864-761-35244	Sequence 35244, A
23	6	14.6	83	17	US-10-425-115-295263	Sequence 295263,
24	6	14.6	87	16	US-10-767-701-54325	Sequence 54325, A
25	6	14.6	95	17	US-10-425-115-341629	Sequence 341629,
26	6	14.6	97	16	US-10-437-963-180730	Sequence 180730,
27	6	14.6	98	17	US-10-425-115-332775	Sequence 332775,
28	6	14.6	100	16	US-10-437-963-137984	Sequence 137984,
29	6	14.6	103	15	US-10-424-599-205861	Sequence 205861,
30	6	14.6	105	15	US-10-424-599-271306	Sequence 271306,
31	6	14.6	105	16	US-10-437-963-141966	Sequence 141966,
32	6	14.6	112	17	US-10-425-115-219369	Sequence 219369,
33	6	14.6	114	16	US-10-437-963-112442	Sequence 112442,
34	6	14.6	117	17	US-10-425-115-266857	Sequence 266857,
35	6	14.6	118	16	US-10-437-963-148111	Sequence 148111,
36	6	14.6	119	17	US-10-425-115-259301	Sequence 259301,
37	6	14.6	122	17	US-10-425-115-294217	Sequence 294217,
38	6	14.6	128	17	US-10-425-115-248863	Sequence 248863,
39	6	14.6	138	15	US-10-424-599-280443	Sequence 280443,
40	6	14.6	140	15	US-10-425-114-41418	Sequence 41418, A
41	6	14.6	140	16	US-10-437-963-153916	Sequence 153916,
42	6	14.6	141	14	US-10-117-087-6	Sequence 4, Appl
43	6	14.6	141	14	US-10-117-087-6	Sequence 6, Appl
44	6	14.6	141	16	US-10-437-963-102617	Sequence 102617,
45	6	14.6	144	16	US-10-437-963-145920	Sequence 145920,

ALIGNMENTS

RESULT 1

US-10-128-558-167

Sequence 167, Application US/10128558

Publication No. US20040219521A1

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom

APPLICANT: Wang, Zhimei

APPLICANT: Weng, Gezhi

APPLICANT: Boyle, Bryan J

APPLICANT: Drmanac, Radjef T

TITLE OF INVENTION: Novel Nucleic Acids and

TITLE OF INVENTION: Polypeptides

FILE REFERENCE: 812A

CURRENT APPLICATION NUMBER: US/10/128, 558

CURRENT FILING DATE: 2002-04-22

PRIOR APPLICATION NUMBER: US 60/339, 453

PRIOR FILING DATE: 2001-12-11

PRIOR APPLICATION NUMBER: US 09/488, 725

PRIOR FILING DATE: 2000-01-21

PRIOR APPLICATION NUMBER: US 09/552, 317

PRIOR FILING DATE: 2000-04-25

PRIOR APPLICATION NUMBER: PCT/US00/35017

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: US 09/491, 404

PRIOR FILING DATE: 2000-01-25

PRIOR APPLICATION NUMBER: PCT/US01/02623

PRIOR FILING DATE: 2001-01-25

PRIOR APPLICATION NUMBER: US 09/496, 914

PRIOR FILING DATE: 2000-02-03

PRIOR APPLICATION NUMBER: US 09/560, 875

PRIOR FILING DATE: 2000-04-27

PRIOR APPLICATION NUMBER: PCT/US01/03800

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 09/515, 126

PRIOR FILING DATE: 2000-02-28

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 412
SOFTWARE: pc_fl_genes version 6.0
SEQ ID NO: 167
LENGTH: 1149
TYPE: PRT
ORGANISM: Homo sapiens
US-10-128-558-167

Query Match 48.8%; Score 20; DB 17; Length 1149;
Best Local Similarity 100.0%; Pred. No. 3.2e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NCLYKIFLFDTHQHSKERA 20
DB 1002 NCLYKIFLFDTHQHSKERA 1021

RESULT 2
US-10-424-599-275868
Sequence 275868, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J
APPLICANT: Kovalic, David K
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 275868
LENGTH: 53
TYPE: PRT
ORGANISM: Glycine max

FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_91128C.1.pdp
US-10-424-599-275868

Query Match 17.1%; Score 7; DB 15; Length 53;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 DTQHRIS 16
DB 32 DTQHRIS 38

RESULT 3
US-10-767-701-41775
Sequence 41775, Application US/10767701
Publication No. US20040172684A1
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/767,701
CURRENT FILING DATE: 2004-01-29
NUMBER OF SEQ ID NOS: 63128
SEQ ID NO 41775
LENGTH: 238
TYPE: PRT
ORGANISM: Sorghum bicolor

FEATURE:
OTHER INFORMATION: Clone ID: SORBI-28MAY03-C6863_1.pdp
US-10-767-701-41775

Query Match 17.1%; Score 7; DB 16; Length 238;
Best Local Similarity 100.0%; Pred. No. 36;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 LGGEKPF 41
DB 211 LGGEKPF 217

RESULT 4
US-10-437-963-187132
Sequence 187132, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei

APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping

TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 187132
LENGTH: 314
TYPE: PRT
ORGANISM: Oryza sativa

FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(314)
OTHER INFORMATION: unsure at all Xaa locations
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_83866C.1.pdp
US-10-437-963-187132

Query Match 17.1%; Score 7; DB 16; Length 314;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 SKERAGA 22
DB 104 SKERAGA 110

RESULT 5
US-10-424-599-149524
Sequence 149524, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J
APPLICANT: Kovalic, David K
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 149524
LENGTH: 349
TYPE: PRT
ORGANISM: Glycine max

FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(349)
OTHER INFORMATION: unsure at all Xaa locations
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_106042C.1.pdp
US-10-424-599-149524

Query Match 17.1%; Score 7; DB 16; Length 314;
Best Local Similarity 100.0%; Pred. No. 46;

Query Match 17.1%; Score 7; DB 15; Length 349;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 PLVDQLE 32
|||||
Db 290 PLVDQLE 296

RESULT 6
US-10-425-114-53502
; Sequence 53502, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 53502
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 701129526_FLI.dep
US-10-425-114-53502

Query Match 17.1%; Score 7; DB 15; Length 406;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 PLVDQLE 32
|||||
Db 335 PLVDQLE 341

RESULT 7
US-10-437-963-110339
; Sequence 110339, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 110339
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_14411C.1.dep
US-10-437-963-110339

Query Match 17.1%; Score 7; DB 16; Length 633;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 PLVDQLE 32
|||||
Db 559 PLVDQLE 565

RESULT 8
US-10-437-963-151847
; Sequence 151847, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 151847
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_51953C.1.dep
US-10-437-963-151847

Query Match 17.1%; Score 7; DB 16; Length 660;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 PLVDQLE 32
|||||
Db 588 PLVDQLE 594

RESULT 9
US-10-437-963-110342
; Sequence 110342, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 110342
; LENGTH: 662
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_14414C.1.dep
US-10-437-963-110342

Query Match 17.1%; Score 7; DB 16; Length 662;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 PLYDQLE 32
|||||
Db 588 PLYDQLE 594

RESULT 10
US-10-424-599-149522
; Sequence 149522, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovacic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 149522
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(665)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PRT_MRT3847_106040C.1.pep
US-10-424-599-149522

Query Match 17.1%; Score 7; DB 15; Length 665;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 PLYDQLE 32
|||||
Db 594 PLYDQLE 600

RESULT 11
US-09-864-761-44493
; Sequence 44493, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmiga-X-1
; CURRENT FILING DATE: 2001-05-23
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669

;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 44493
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005229.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.86
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.86
US-09-864-761-44493

Query Match 14.6%; Score 6; DB 9; Length 36;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 GGEKXF 41
|||||
Db 4 GGEKXF 9

RESULT 12
US-10-425-115-340619
; Sequence 340619, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovacic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT FILING DATE: 2003-04-28
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 340619
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_73815C.1.pep
US-10-425-115-340619

Query Match 14.6%; Score 6; DB 17; Length 45;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LYKIFL 8
|||||

```
Db          14 LYKIFL 19

RESULT 13
; Sequence 197, Application US/09989920
; Patent No. US20020172957A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sel-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Pro
; FILE REFERENCE: DEX-0291
; CURRENT APPLICATION NUMBER: US/09/989,920
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/252,500
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 284
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 197
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-989-920-197

Query Match          14.6%; Score 6; DB 9; Length 48;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          26 PLVDQL 31
          |||||
Db          42 PLVDQL 47

RESULT 14
US-10-424-599-219715
; Sequence 219715, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 219715
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURES:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_40430C.1.pep
US-10-424-599-219715

Query Match          14.6%; Score 6; DB 15; Length 51;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          31 LENRLG 36
          |||||
Db          37 LENRLG 42

RESULT 15
US-10-029-386-29247
; Sequence 29247, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: A60MCA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 29247
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURES:
; OTHER INFORMATION: MAP TO CHR17.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
; OTHER INFORMATION: SWISSPROT HIT: P52631, EVALUO 1.00e-21
US-10-029-386-29247

Query Match          14.6%; Score 6; DB 14; Length 52;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          14 FISKER 19
          |||||
Db          9 FISKER 14

Search completed: February 1, 2005, 15:44:52
Job time : 61.7778 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:55:01 / Search time 13.5556 Seconds
(without alignments)
291.016 Million cell updates/sec

Title: SEQ5GLY2017

Perfect score: 41
Sequence: 1 NCLYKIFLPDQHFISKRA.....ALMPLVDQLENRLGSEK 41

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size: 0

Total number of hits satisfying chosen parameters: 279530

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

1: PIR_79:.*
2: p1r1:.*
3: p1r2:.*
4: p1r3:.*
5: p1r4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	17.1	232	2	JC7706
2	7	17.1	276	2	AG3305
3	7	17.1	584	2	S77647
4	7	17.1	651	2	T06979
5	6	14.6	38	2	A21851
6	6	14.6	154	2	D39384
7	6	14.6	165	2	B70385
8	6	14.6	184	2	AE1538
9	6	14.6	196	2	G65039
10	6	14.6	196	2	F90919
11	6	14.6	196	2	C85768
12	6	14.6	196	2	D64918
13	6	14.6	204	2	A59263
14	6	14.6	222	2	A59263
15	6	14.6	226	1	UJ0221
16	6	14.6	226	1	S39262
17	6	14.6	226	1	I49589
18	6	14.6	228	1	A40402
19	6	14.6	228	1	A42929
20	6	14.6	229	2	T08040
21	6	14.6	235	2	B41845
22	6	14.6	239	2	AE0883
23	6	14.6	245	2	E90239
24	6	14.6	251	2	G84036
25	6	14.6	251	2	C83083
26	6	14.6	256	2	E72257
27	6	14.6	264	2	G83165
28	6	14.6	268	2	D71325
29	6	14.6	269	2	AD2833

30	6	14.6	283	2	H97610	hypothetical prote
31	6	14.6	295	2	D90252	conserved hypothet
32	6	14.6	302	2	T08522	NTP-binding protei
33	6	14.6	302	2	S32178	NTP-binding protei
34	6	14.6	302	2	S70151	tnac protein homol
35	6	14.6	317	2	G83593	glutathione synthe
36	6	14.6	326	2	AE1169	hypothetical prote
37	6	14.6	326	2	AG1526	hypothetical prote
38	6	14.6	332	2	G95380	probable periplasm
39	6	14.6	335	2	T50935	isomerase/decarbox
40	6	14.6	347	2	AF2645	flagellar motor sw
41	6	14.6	347	2	F97427	flagellar motor sw
42	6	14.6	350	2	C91014	hypothetical prote
43	6	14.6	350	2	E85858	probable subunit o
44	6	14.6	350	2	HE4988	cytochrome c-type
45	6	14.6	351	2	H70619	hypothetical prote

ALIGNMENTS

RESULT 1

JC7706 beta crystallin B1 protein - zebra fish

C:Species: Brachydanio rerio (zebra fish)

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Jul-2004

C:Accession: JC7706

R:Chen, J.Y.; Chang, B.E.; Chen, Y.H.; Lin, C.J.F.; Wu, J.L.; Kuo, C.M.

Biochem. Biophys. Res. Commun. 285, 105-110, 2001

A:Title: Molecular cloning, developmental expression, and hormonal regulation of zebrafi

A:Reference number: JC7706; PMID:21331298; PMID:11437379

A:Accession: JC7706

A:Molecule type: mRNA

A:Residues: 1-232 <CH>

A:Cross-references: UNIPROT:Q90WT1; GB:AJ317957

C:Comment: This protein is involved in regulation by growth factors.

C:Superfamily: beta-crystallin

F:20-25/Region: active proline- and alanine-rich motif #status predicted

Query Match 17.1%; Score 7; DB 2; Length 232;

Best Local Similarity 100.0%; Pred. No. 6.3;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 YKIFLPD 10

Db 41 YKIFLPD 47

RESULT 2

AG3305 23S ribosomal RNA methyltransferase (EC 2.1.1.-) [imported] - Brucella melitensis (strai

C:Species: Brucella melitensis

C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004

C:Accession: AG3305

R:DelVecchio, V.G.; Kaput, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis

A:Reference number: AD3552; PMID:11756688

A:Accession: AG3305

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-276 <KIR>

A:Cross-references: UNIPROT:Q8Y1L5; UNIPROT:Q8F2A1; GB:AE008917; PDB:AAU51610.1; PDB:91

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BMEI0429

A:Map position: I

C:Keywords: methyltransferase

Query Match 17.1%; Score 7; DB 2; Length 276;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 LENR1GG 37
 DB 67 LENR1GG 73

RESULT 3

S77647
 mobilization protein A - Lactococcus lactis self-transfer sex factor
 C/Species: Lactococcus lactis
 C/Date: 24-Oct-1998 #sequence_revision 24-Oct-1998 #text_change 21-Jul-2000
 C/Accession: S77647
 R/Shearman, C.A.; Godon, J.J.; Gasson, M.
 M01. Microbiol. 21, 45-53, 1996
 A/Title: Splicing of a group II intron in a functional transfer gene of Lactococcus lactis
 A/Reference number: S77646; MUID:97000348; PMID:8643433
 A/Accession: S77647
 A/Molecule type: DNA
 A/Residues: 1-584 <SHE>
 A/Cross-references: EMBL:X89922; NID:G1296826; PIDD:CAAG1995.1; PID:G1296828
 A/Experimental source: strain MG1363
 C/Genetics:
 A/Gene: mobA
 A/Mobile element: self-transfer sex factor
 A/Intons: 171/3
 C/Function:
 A/Description: involved in conjugation, probably by introducing a single-stranded nick
 C/Superfamily: Lactococcus lactis self-transfer sex factor mobilization protein A
 C/Keywords: conjugation; replication

Query Match 17.1%; Score 7; DB 2; Length 584;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 VDQLENR 34
 DB 396 VDQLENR 402

RESULT 4

T06979
 polyadenylate-binding protein - wheat
 N/Alternate names: poly(A)-binding protein
 C/Species: Triticum aestivum (common wheat)
 C/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004
 C/Accession: T06979
 R/Le, H.; Chang, S.C.; Tanguay, R.L.; Gallie, D.R.
 Eur. J. Biochem. 243, 350-357, 1997
 A/Title: The wheat poly(A)-binding protein functionally complements pab1 in yeast.
 A/Reference number: Z12044; MUID:97182620; PMID:9030755
 A/Accession: T06979
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-651 <LEH>
 A/Cross-references: UNIPROT:P9316; EMBL:U81318; NID:G1737491; PIDD:AAB38974.1; PID:G1737491
 A/Experimental source: sprout tips from 5 day old growing sprouts
 C/Genetics:
 A/Note: wheatpb

C/Superfamily: polyadenylate-binding protein; ribonucleoprotein repeat homology
 C/Keywords: nucleus; RNA binding
 F/33-100/Domain: ribonucleoprotein repeat homology <RRM1>
 F/121-187/Domain: ribonucleoprotein repeat homology <RRM2>
 F/212-278/Domain: ribonucleoprotein repeat homology <RRM3>
 F/315-381/Domain: ribonucleoprotein repeat homology <RRM4>

Query Match 17.1%; Score 7; DB 2; Length 651;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 PLYDQLE 32
 DB 578 PLYDQLE 584

RESULT 5

A21851
 22k factor - human (fragment)
 C/Species: Homo sapiens (man)
 C/Date: 28-Sep-1987 #sequence_revision 28-Sep-1987 #text_change 09-Jul-2004
 C/Accession: A21851
 R/Damme, J.V.; De Ley, M.; Opdenacker, G.; Billiau, A.; De Smer, P.
 Nature 314, 266-268, 1985
 A/Title: Homogeneous interferon-inducing 22k factor is related to endogenous pyrogen and
 A/Reference number: A21851; MUID:85163727; PMID:3970526
 A/Accession: A21851
 A/Status: preliminary
 A/Molecule type: protein
 A/Residues: 1-38 <DAM>
 A/Cross-references: UNIPROT:Q7M4S7
 C/Superfamily: interleukin-1

Query Match 14.6%; Score 6; DB 2; Length 38;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LFDPTQH 13
 DB 10 LFDPTQH 15

RESULT 6

D39384
 finger protein HTP6 - human (fragment)
 C/Species: Homo sapiens (man)
 C/Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 08-Dec-2000
 C/Accession: D39384
 R/Bellefroid, E.J.; Poncelet, D.A.; Lecocq, P.J.; Revelant, O.; Martial, J.A.
 Proc. Natl. Acad. Sci. U.S.A. 89, 3608-3612, 1991
 A/Title: The evolutionarily conserved Knueppel-associated box domain defines a subfamily
 A/Reference number: A39384; MUID:91219421; PMID:2023909
 A/Accession: D39384
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-154 <BEL>

A/Cross-references: GB:M61869; NID:G184449; PID:G184450
 C/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
 C/Keywords: DNA binding; zinc finger

Query Match 14.6%; Score 6; DB 2; Length 154;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 KIFLFD 10
 DB 102 KIFLFD 107

RESULT 7

E70385
 conserved hypothetical protein aq_987 - Aquifex aeolicus
 C/Species: Aquifex aeolicus
 C/Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
 C/Accession: E70385
 R/Beckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
 V.
 Nature 392, 353-358, 1998
 A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A/Reference number: A70300; MUID:98196666; PMID:9537320
 A/Accession: E70385
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-165 <AQF>

A/Cross-references: UNIPROT:O67112; GB:AE000717; NID:G2983492; PIDD:AA07079.1; PID:G2983492
 A/Experimental source: strain VFS
 C/Genetics:
 A/Gene: aq_987

C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0445

Query Match 14.6%; Score 6; DB 2; Length 165;

Best Local Similarity 100.0%; Pred. No. 54;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 RLGSE 39

|||||

38 RLGSE 43

RESULT 8

AE1538

transcription regulator TetR/AcrR family homolog lin0845 [imported] - Listeria innocua

C:Species: Listeria innocua

C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004

C:Accession: AE1538

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H.

D.; Jones, L.M.; Karet, U.

Science 294, 849-852, 2001

A:Authors: Kreft, U.; Kunz, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schlueter, T.; Simoes, N.; Tlertre, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A.; Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; PMID:11679669

A:Accession: AE1538

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-184 <GLA>

A:Cross-references: UNIPROT:Q92D68; GB:AE592022; PIDN:CAC96077.1; PID:g1641326; GSPDB:G

A:Experimental source: strain Clp11262

C:Genetics:

A:Gene: lin0845

Query Match 14.6%; Score 6; DB 2; Length 184;

Best Local Similarity 100.0%; Pred. No. 60;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 LGSEK 40

|||||

73 LGSEK 78

RESULT 9

G65039

hypothetical protein b2612 - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002

C:Accession: G65039

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; PMID:97426617; PMID:9278503

A:Accession: G65039

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-196 <BLAT>

A:Cross-references: GB:AE000347; GB:U00096; NID:g2367142; PIDN:AACT5661.1; PID:g1788965;

A:Experimental source: strain K-12, substrain MG1655

Query Match 14.6%; Score 6; DB 2; Length 196;

Best Local Similarity 100.0%; Pred. No. 63;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 LMPPLV 28

|||||

130 LMPPLV 135

RESULT 10

P90919

repressor for uid operon [imported] - Escherichia coli (strain O157:H7, substrain RMD 0

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004

C:Accession: P90919

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasavar, N.; Yasunaga, T.; Kunara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno

A:Reference number: A99629; PMID:21156231; PMID:11258796

A:Accession: P90919

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-196 <HAV>

A:Cross-references: UNIPROT:Q59431; GB:BA000007; PIDN:BA835749.1; PID:g13361793; GSPDB:G

A:Experimental source: strain O157:H7, substrain RMD 0509952

C:Genetics:

A:Gene: ECe2326

Query Match 14.6%; Score 6; DB 2; Length 196;

Best Local Similarity 100.0%; Pred. No. 63;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 HFISKE 18

|||||

51 HFISKE 56

RESULT 11

C85768

repressor for uid operon [imported] - Escherichia coli (strain O157:H7, substrain EDL933

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004

C:Accession: C85768

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

iller, U.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamouels, K.; Apodaca,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; PMID:21074935; PMID:11206551

A:Accession: C85768

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-196 <STO>

A:Cross-references: UNIPROT:Q59431; GB:AE005174; NID:g12515604; PIDN:AA656607.1; GSPDB:G

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: uidR

Query Match 14.6%; Score 6; DB 2; Length 196;

Best Local Similarity 100.0%; Pred. No. 63;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 HFISKE 18

|||||

51 HFISKE 56

RESULT 12

D64918

glucuronide repressor guer - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004

C:Accession: D64918

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; PMID:97426617; PMID:9278503

A:Accession: D64918

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-196 <BLAT>

A:Cross-references: UNIPROT:Q59431; GB:AE000257; GB:U00096; NID:g1787898; PIDN:AACT74690.

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: guar; uidr
 C:Function:
 A:Description: repressor of uidRABC (= gusRABC) operon
 C:Keywords: transcription regulation

Query Match 14.6%; Score 6; DB 2; Length 196;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 HFSKE 18
 |||||
 Db 51 HFSKE 56

RESULT 13

Conserved hypothetical protein CC0375 [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
 C:Accession: F87295
 R:NIEMAN, W.C.; FELDBLYUM, T.V.; PAULSEN, I.T.; NELSON, K.E.; EISEN, J.; HEIDELBERG, J.
 B.; LAUB, M.T.; DEBOY, R.T.; DODSON, R.J.; DURKIN, A.S.; GWINN, M.L.; HAFU, D.H.; KOLOD
 PROC. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A>Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: AB7249; MUID:21173698; PMID:11259647
 A:Accession: F87295
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-204 <STO>
 A:Cross-references: UNIPROT:Q9AB57; GB:AE005673; NID:g13421530; PIDN:MAK22362.1; GSPDB:G
 C:Genetics:
 A:Gene: CC0375

Query Match 14.6%; Score 6; DB 2; Length 204;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 RIGGE 39
 |||||
 Db 181 RIGGE 186

RESULT 14

tetraspan TSPAN-2 - human
 C:Species: Homo sapiens (man)
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
 C:Accession: A59263
 R:Todd, S.C.; Doctor, V.S.; Levy, S.
 Biochim. Biophys. Acta 139, 101-104, 1998
 A>Title: Sequences and expression of six new members of the tetraspanin/TW4SF family.
 A:Reference number: A59258; MUID:98390278; PMID:9714763
 A:Accession: A59263
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-222 <TOD>
 A:Cross-references: UNIPROT:O60636; GB:AF054839; NID:g2997742; PIDN:AAC69715.1; PID:g299
 C:Genetics:
 A:Gene: TSPAN-2
 C:Superfamily: CD9 antigen

Query Match 14.6%; Score 6; DB 2; Length 222;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 AGALM 25
 |||||
 Db 65 AGALM 70

RESULT 15

JX0221

CD9 antigen - bovine
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 09-Oct-1992 #sequence_revision 09-Aug-1996 #text_change 09-Jul-2004
 C:Accession: JX0221

R:Martin-Alonso, J.M.; Hernandez, N.; Ghosh, S.; Coca-Prados, M.
 J. Biochem. 112, 63-67, 1992
 A>Title: Molecular cloning of the bovine CD9 antigen from ocular ciliary epithelial cells.
 A:Reference number: JX0221; MUID:93054422; PMID:1339429
 A:Accession: JX0221
 A:Molecule type: mRNA
 A:Residues: 1-226 <MAR>

A:Cross-references: UNIPROT:P30932; GB:M81720; NID:g162820; PIDN:AAA30439.1; PID:g162821
 A:Experimental source: ocular ciliary epithelial cell
 C:Superfamily: CD9 antigen

C:Keywords: glycoprotein; transmembrane protein
 F:2-226/Product: CD9 antigen #status predicted <MAT>
 F:12-11/Domain: intracellular #status predicted <CV1>
 F:12-35/Domain: transmembrane #status predicted <TM1>
 F:36-53/Domain: extracellular #status predicted <EX1>
 F:54-76/Domain: transmembrane #status predicted <TM2>
 F:77-80/Domain: intracellular #status predicted <CY2>
 F:81-109/Domain: transmembrane #status predicted <TM3>
 F:110-192/Domain: extracellular #status predicted <EX2>
 F:193-219/Domain: transmembrane #status predicted <TM4>
 F:220-226/Domain: intracellular #status predicted <CY3>
 F:50/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 14.6%; Score 6; DB 1; Length 226;
 Best Local Similarity 100.0%; Pred. No. 71;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 AGALM 25
 |||||
 Db 64 AGALM 69

Search completed: February 1, 2005, 15:33:06
 Job time: 15.5556 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:54:06 ; Search time 70.5556 Seconds
(without alignment)
334.352 Million cell updates/sec

Title: SEQ5GLY2017
Perfect score: 41
Sequence: 1 NCLYKIFLPTQHPISKERA.....ALMPLVDQLENRLGSEK 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1825181 seqs, 575374646 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1810864

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Uniprot_02.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	41	100.0	Q96ES5	Q96ES5 homo sapien
2	41	100.0	Q6P197	Q6P197 homo sapien
3	41	100.0	AAH65205	AAH65205 homo sapi
4	31	75.6	Q8N7L7	Q8N7L7 homo sapien
5	20	48.8	BP28_MACPA	BP28_MACPA
6	20	48.8	2144_1	BP28_HUMAN
7	18	43.9	Q8VCK1	Q8VCK1 mus musculu
8	18	43.9	Q8CCT5	Q8CCT5 mus musculu
9	18	43.9	Q7T153	Q7T153 brachydanio
10	8	19.5	Q7T152	Q7T152 brachydanio
11	8	19.5	Q7S148	Q7S148 brachydanio
12	8	19.5	2159_2	Q7S148
13	12	17.1	EX7S_VIBPA	EX7S_VIBPA
14	14	17.1	EX7S_VIBVU	EX7S_VIBVU
15	15	17.1	Q7M47	Q7M47 vibrio vuln
16	15	17.1	Q9W11	Q9W11 brachydanio
17	16	17.1	Q6D328	Q6D328 brachydanio
18	17	17.1	Q98H1	Q98H1 rhizobium 1
19	18	17.1	Q6UN5	Q6UN5 oryza sativ
20	19	17.1	AAO56324	AAO56324 oryza sat
21	20	17.1	Q8FZ1	Q8FZ1 bruceella su
22	22	17.1	Q8Y115	Q8Y115 bruceella me
23	23	17.1	Q6M213	Q6M213 methylocyst
24	24	17.1	CAE48356	CAE48356 methyloc
25	25	17.1	Q7MA16	Q7MA16 wolinnella s
26	26	17.1	Q8DKO2	Q8DKO2 synecococc
27	27	17.1	Q9M6S4	Q9M6S4 nicotiana t
28	28	17.1	Q8KY0	Q8KY0 uncultured
29	29	17.1	Q8722	Q8722 lactococcus
30	30	17.1	LTRB_LACIA	LTRB_LACIA
31	31	17.1	P93616	P93616 triticum ae
32	32	17.1	Q86924	Q86924 dictyosteli

ALIGNMENTS

RESULT 1	ID	Q96ES5	PRELIMINARY	PRT	349 AA
AC	Q96ES5				
DT	01-DEC-2001	(TREMBLrel. 19, Created)			
DT	01-DEC-2001	(TREMBLrel. 19, Last sequence update)			
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)			
DE	FLJ10359	protein.			
GN	Name=FLJ10359				
OS	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_Taxid=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Ovary;				
RX	MEDLINE=22388257; PubMed=12477932;				
RA	Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,				
RA	Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,				
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,				
RA	Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh P.,				
RA	Diatchenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,				
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,				
RA	Brownstein M.J., Uediri T.B., Toshitsuki S., Carinici P., Prange C.,				
RA	Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,				
RA	Bozak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,				
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,				
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,				
RA	Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,				
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,				
RA	Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,				
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,				
RA	Krzywinski M.I., Skalska U., Smalov D.E., Scherch A., Schein J.E.,				
RA	Jones S.J., Maira M.A.,				
RT	"Generation and initial analysis of more than 15,000 full-length human				
RT	and mouse cDNA sequences."				
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Ovary;				
RA	Strauberg R.,				
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL, BC011983; AATH1983.1; -				
DR	InterPro, IPR006938; ARM.				
SO	SEQUENCE 349 AA; 39921 MW; 3A359597FF7079EB CRC64;				
Query Match	100.0%; Score 41; DB 2; Length 349;				
Best Local Similarity	100.0%; Pred. No. 1.4e-35;				
Matches	41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
OY	1 NCLYKIFLPTQHPISKERAGALMPLVDQLENRLGSEK 41				
DB	202 NCLYKIFLPTQHPISKERAGALMPLVDQLENRLGSEK 242				

```

RESULT 2
Q6P197 PRELIMINARY; PRT; 1106 AA.
ID Q6P197;
AC Q6P197;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE FUJ10359 protein (Fragment).
GN Name=FUJ10359;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shmamen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.D., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Skimutz J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA Strausberg R.;
RA Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC065205; AAH65205.1; -.
FT NON_TER 1
SQ SEQUENCE 1106 AA; 125359 MW; 09F7CE94042302C4 CRC64;

Query Match 100.0%; Score 41; DB 2; Length 1106;
Best Local Similarity 100.0%; Pred. NO. 3.4e-35;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NCLYKIFLFDTOHFISKERAGALMPLVDQLENRLGGEK 41
Db 959 NCLYKIFLFDTOHFISKERAGALMPLVDQLENRLGGEK 999

RESULT 3
AAH65205 PRELIMINARY; PRT; 1106 AA.
ID AAH65205;
AC AAH65205;
DT 02-MAR-2004 (TREMBlrel. 27, Created)
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
DE FUJ10359 protein (Fragment).
GN Name=FUJ10359;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;

```

```

RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shmamen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.D., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Skimutz J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA Strausberg R.;
RA Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC065205; AAH65205.1; -.
FT NON_TER 1
SQ SEQUENCE 1106 AA; 125359 MW; 09F7CE94042302C4 CRC64;

```

```

Query Match 100.0%; Score 41; DB 2; Length 1106;
Best Local Similarity 100.0%; Pred. NO. 3.4e-35;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 NCLYKIFLFDTOHFISKERAGALMPLVDQLENRLGGEK 41
Db 959 NCLYKIFLFDTOHFISKERAGALMPLVDQLENRLGGEK 999

```

```

RESULT 4
Q8N7L7 PRELIMINARY; PRT; 897 AA.
ID Q8N7L7;
AC Q8N7L7;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein FLJ40893.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA PubMed=14702039;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obaishi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto K., Saito K., Kawai Y., Isono Y., Nakamura N., Nagahata K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoki T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa M., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Niimiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiwasa S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fuji A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,

```

RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Saeki M.,
 RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashta R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
 RT "Complete sequencing and characterization of 21,243 full-length human
 RT cDNAs.";
 RL Nat. Genet. 36:40-45 (2004).
 DR EMBL; AK098212; BAC05261.1; -
 DR InterPro; IPR008938; ARM.
 SQ SEQUENCE 897 AA; 101574 MW; 5FF6A94FB855895 CRC64;

Query Match 75.6%; Score 31; DB 2; Length 897;
 Best Local Similarity 100.0%; Pred. No. 1,7e-24;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NCLYKIFLFTQHFISKERAGALMPLVDOL 31
 DB 750 NCLYKIFLFTQHFISKERAGALMPLVDOL 780

RESULT 5
 BP28 MACFA STANDARD; PRT; 958 AA.
 AC O9GM44;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Protein BAP28 (OmpA-17571) (Fragment).
 GN Name=BAP28;
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 CC Cercopithecinae; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
 RA Suzuki Y., Sugano S., Hashimoto K.;
 RT "Isolation of full-length cDNA clones from macaque brain cDNA
 RT libraries.";
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 CC -1 SIMILARITY: Belongs to the BAP28 family.
 CC -1 SIMILARITY: Contains 1 HEAT repeat.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC cDNAs.";
 RL Nat. Genet. 36:40-45 (2004).
 DR EMBL; AB049842; BAB16728.1; ALT_INIT.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR000357; HEAT.
 DR PROSITE; PS50077; HEAT_REPEAT; FALSE_NEG.
 FT NON_TER 1
 FT REPEAT 920 956 HEAT.
 SQ SEQUENCE 958 AA; 108644 MW; 3DBD95C3623CFB31 CRC64;

Query Match 48.8%; Score 20; DB 1; Length 958;
 Best Local Similarity 100.0%; Pred. No. 1.2e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NCLYKIFLFTQHFISKERAGALMPLVDOL 20
 DB 1 NCLYKIFLFTQHFISKERAGALMPLVDOL 20

DB 811 NCLYKIFLFTQHFISKERAGALMPLVDOL 830

RESULT 6
 BP28 HUMAN STANDARD; PRT; 2144 AA.
 AC O9H583; O9NW23;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 05-JUN-2004 (Rel. 44, Last annotation update)
 DE Protein BAP28.
 GN Name=BAP28;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANTS SER-1694, ALA-1854, ASP-1967 AND
 RP GLY-2017.
 RA Bouguetel L., Chumakov I., Barry C., Cohen-Akenine A.;
 RT "A novel BAP28 gene and protein.";
 RL Patent number WO0100669, 04-JUN-2001.
 RN [2]
 RP SEQUENCE OF 1534-2144 FROM N.A.
 RA Cobley V.;
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1777-2144 FROM N.A.
 EX PubMed14702039; DOI=10.1038/ng1285;
 RA Ota T., Suzuki Y., Nishikawa T., Ohtsuki T., Sugiyama T., Irie R.,
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
 RA Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
 RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
 RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
 RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
 RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
 RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,
 RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hoota T.,
 RA Kusano Y., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
 RA Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,
 RA Imose N., Muesashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura A., Fukuzumi Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Saeki M.,
 RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashta R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
 RT "Complete sequencing and characterization of 21,243 full-length human
 RT cDNAs.";
 RL Nat. Genet. 36:40-45 (2004).
 DR EMBL; AX067150; CAC26776.1; -
 DR EMBL; AL136105; CAC15948.1; -

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC cDNAs.";
 RL Nat. Genet. 36:40-45 (2004).
 DR EMBL; AX067150; CAC26776.1; -
 DR EMBL; AL136105; CAC15948.1; -

```

DR EMBL; AK001221; BAA91564.1; ALT_INIT.
DR SWISS-2DPAGE; Q9H583; HUMAN.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000357; HEAT.
DR PROSITE; PSS0077; HEAT_REPEAT; FALSE_NEG.
KW Polymorphism.
FT REPEAT 2106 2142 HEAT.
FT VARIANT 1694 1694 N->S.
FT VARIANT 1854 1854 /FTID=VAR_010939.
FT VARIANT 1967 1967 V->A.
FT VARIANT 2017 2017 N->D.
FT VARIANT 2017 2017 /FTID=VAR_010941.
FT VARIANT 2017 2017 E->G.
FT SEQUENCE 2144 AA; 242355 MW; D66816E78D8C9B7 CRC64;

Query Match 48.8%; Score 20; DB 1; Length 2144;
Best Local Similarity 100.0%; Pred. No. 2,3e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NCLYKIFLPTQHFISKERA 20.
Db 1997 NCLYKIFLPTQHFISKERA 2016

RESULT 7
Q8VCX1 PRELIMINARY; PRT; 349 AA.
AC Q8VCX1;
DT 01-MAR-2002 (TRENBLREL. 20, Created)
DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)
DE 01-MAR-2004 (TRENBLREL. 26, Last annotation update)
DE BC019693 protein.
GN Name=BC019693;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=mx FVB/N;
RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shermen C.M., Schuler G.D.,
RA Altchul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.,
RA Krzyzinski M.I., Skalska U., Smaltus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Maira M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=mx FVB/N;
RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RA Strussberg R.;
RA Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC019693; AAH19693.1; -.
DR MGD; MGI:2384983; BC019693.
DR InterPro; IPR008938; ARM.

```

```

SQ SEQUENCE 349 AA; 40166 MW; 9763D0331AD0F515 CRC64;

Query Match 43.9%; Score 18; DB 2; Length 349;
Best Local Similarity 100.0%; Pred. No. 7,9e-11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 ALMPVLDQLENRLGEE 39
Db 223 ALMPVLDQLENRLGEE 240

RESULT 8
Q8CCT5 PRELIMINARY; PRT; 408 AA.
AC Q8CCT5;
DT 01-MAR-2003 (TRENBLREL. 23, Created)
DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)
DT 01-MAR-2004 (TRENBLREL. 26, Last annotation update)
DE Mus musculus adult male olfactory brain cDNA, RIKEN full-length
DE enriched library, clone:643040D06 product:hypothetical ARM repeat
DE structure containing protein, full insert sequence.
GN Name=BC019693;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RA "High-efficiency full-length cDNA cloning.";
RL Mech. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RA The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugihara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Kono H., Akiyama J., Nishi K., Kitsuai T., Taishiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Wachihi M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hirata T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Koyama M., Koyama S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishii K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Saeki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK032134; BAC27721.1; -.
DR MGI; MGI:384983; BC019693.
DR InterPro; IPR008938; ARM.
KM Hypothetical protein.
SQ
SEQUENCE 408 AA; 4668 MW; 42AB9EB13CA3FE67 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 9e-11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 ALMPPLVDQLENRLGCEE 39
DB 282 ALMPPLVDQLENRLGCEE 299

RESULT 9
OY 07T153 PRELIMINARY; PRT; 1278 AA.
AC 07T153;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE S1:ZC146P4.2.2 (Novel protein similar to human BAP28) (Fragment).
GN Name=S1:ZC146P4.2;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Garner P.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL732629; CAEL7602.1; -.
FT NON TER 1
SQ SEQUENCE 1278 AA; 143523 MW; E6C9FC81B77EB1A9 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 PLVDQLEN 33
DB 1214 PLVDQLEN 1221

RESULT 10
OY 07T152 PRELIMINARY; PRT; 1336 AA.
AC 07T152;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE S1:ZC146P4.2.1 (Novel protein similar to human BAP28) (Fragment).
GN Name=S1:ZC146P4.2;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RA Garner P.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL732629; CAEL7603.1; -.
FT NON TER 1
SQ SEQUENCE 1336 AA; 150326 MW; AA949557F21ACBCE CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 PLVDQLEN 33
DB 1214 PLVDQLEN 1221

RESULT 11
OY 07SY48 PRELIMINARY; PRT; 2159 AA.
AC 07SY48;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein zgc:63510.
GN Name=zgc:63510;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Whole body;
RX MEDLINE=22388257; PubMed=12477932.
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares W.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshyuk S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Scherch A., Schein J.E.,
RA Jones S.J., Maitra M.A.;
OS "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Whole body;
RA Strausberg R.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC055128; AAH55128.1; -.
KM Hypothetical protein.
SQ SEQUENCE 2159 AA; 242048 MW; F00DBBA1D597E70B CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 PLVDQLEN 33
DB 2037 PLVDQLEN 2044

```

```

RESULT 12
EX7S_VIBPA STANDARD; PRT; 80 AA.
ID EX7S_VIBPA
AC 08RT8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable exodeoxyribonuclease VII small subunit (EC 3.1.11.6)
DE (Exonuclease VII small subunit).
GN Name=xseB; OrderedLocNames=VP0688;
OS Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
NCBI_TaxID=670;
OX
RN
RP SEQUENCE FROM N.A.
RC STRAIN=RIMD 2210633 / Serotype O3:k6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Nishimura M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae.";
RL Lancet 361:743-749(2003).
CC -1- FUNCTION: Bidirectionally degrades single-stranded DNA into large
CC acid-insoluble oligonucleotides, which are then degraded further
CC into small acid-soluble oligonucleotides (By similarity).
CC -1- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5'- to 3'-
CC or 3'- to 5'-direction to yield nucleoside 5'-phosphates.
CC -1- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the xseB family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF005075; BAC56951.1; -.
DR HAMAP; MF_00337; -; 1.
DR InterPro; IPR003761; Exonuc_VII_S.
DR Pfam; PF02609; Exonuc_VII_S; 1.
DR Complete proteome; Exonuclease; Hydrolase; Nuclease.
SQ SEQUENCE 80 AA; 8879 MW; 03B881D68BEC4F6A CRC64;

Query Match 17.1%; Score 7; DB 1; Length 80;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 LVDQLEN 33
Db 21 LVDQLEN 27

```

```

RESULT 13
EX7S_VIBVU STANDARD; PRT; 80 AA.
ID EX7S_VIBVU
AC 08DPA5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable exodeoxyribonuclease VII small subunit (EC 3.1.11.6)
DE (Exonuclease VII small subunit).
GN Name=xseB; OrderedLocNames=VVI0313;
OS Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
NCBI_TaxID=672;
OX
RN

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee Y.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Bidirectionally degrades single-stranded DNA into large
CC acid-insoluble oligonucleotides, which are then degraded further
CC into small acid-soluble oligonucleotides (By similarity).
CC -1- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5'- to 3'-
CC or 3'- to 5'-direction to yield nucleoside 5'-phosphates.
CC -1- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the xseB family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE016798; AAC08843.1; -.
DR HAMAP; MF_00337; -; 1.
DR InterPro; IPR003761; Exonuc_VII_S.
DR Pfam; PF02609; Exonuc_VII_S; 1.
DR TIGRPFAMs; TIGR01280; xseB; 1.
DR Complete proteome; Exonuclease; Hydrolase; Nuclease.
SQ SEQUENCE 80 AA; 8756 MW; FE23B015BD49F5ED CRC64;

Query Match 17.1%; Score 7; DB 1; Length 80;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 LVDQLEN 33
Db 21 LVDQLEN 27

```

```

RESULT 14
Q7MN47 PRELIMINARY; PRT; 80 AA.
ID Q7MN47
AC Q7MN47;
DT 01-MAR-2004 (TrEMBL;rel. 26, Created)
DT 01-MAR-2004 (TrEMBL;rel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBL;rel. 26, Last annotation update)
DE Exonuclease VII small subunit.
GN Name=VVO870;
OS Vibrio vulnificus (strain VVO16).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
NCBI_TaxID=196600;
OX
RN
RP SEQUENCE FROM N.A.
RX PubMed=14656965;
RA Chen C.-Y., Wu K.-M., Chang Y.-C., Chang C.-H., Tsai H.-C.,
RA Liao T.-L., Liu Y.-M., Chen H.-J., Shen A.B.-T., Li J.-C., Su T.-L.,
RA Shao C.-P., Lee C.-T., Hor L.-I., Tsai S.-F.;
RT "Comparative genome analysis of Vibrio vulnificus, a marine
RT pathogen.";
RL Genome Res. 13:2577-2587(2003).
DR EMBL; AP005333; BAC93634.1; -.
DR GO; GO:0009318; C:exodeoxyribonuclease VII complex; IEA.
DR GO; GO:0008855; F:exodeoxyribonuclease VII activity; IEA.
DR GO; GO:0004527; F:exonuclease activity; IEA.
DR GO; GO:006508; P:DNA catabolism; IEA.
DR InterPro; IPR003761; Exonuc_VII_S.
DR Pfam; PF02609; Exonuc_VII_S; 1.
DR ProDom; PD028235; Exonuc_VII_S; 1.
DR TIGRPFAMs; TIGR01280; xseB; 1.
DR Exonuclease.
KW

```

SEQ SEQUENCE 80 AA; 8756 MW; FE23B015BD49F5ED CRC64;

Query Match 17.1%; Score 7; DB 2; Length 80;

Best Local Similarity 100.0%; Pred. No. 18;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 27 LVDQLEN 33

DB 21 LVDQLEN 27

RESULT 15

Q90WT1 PRELIMINARY; PRT; 232 AA.
 ID Q90WT1
 AC Q90WT1;
 DT 01-DEC-2001 (TRENBLREL. 19, Created)
 DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)
 DT 01-MAR-2004 (TRENBLREL. 26, Last annotation update)
 DE Crystallin B1 protein.
 GN Name=crystb1; Synonyms=crystallin B1;
 OS Brachydanio rerio (zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI TaxId=7955;
 [1]
 RP SEQUENCE FROM N.A.
 RA Chen J.Y.;
 RL Theis (2001), Department of Institute of Zoology, Academia Sinica,
 Taipei, Taiwan.
 CC -!- FUNCTION: Crystallins are the dominant structural components of
 the vertebrate eye lens (By similarity).
 CC -!- DOMAIN: Has a two-domain beta-structure, folded into four very
 similar Greek key motifs (by similarity).
 CC -!- SIMILARITY: Belongs to the beta/gamma-crystallin family.
 CC EMBL; AJ317957; CAC84899.1; -.
 DR PIR; JC7706; JC7706.
 DR HSSP; P26775; 1BD7.
 DR ZFIN; ZDB-GENE-010813-1; crystb1.
 DR InterPro; IPR01064; Crystallin.
 DR InterPro; IPR011024; G-crystallin_SF.
 DR Pfam; PF00030; Crystall; 2.
 DR PRINTS; PR01367; BGCRCRYSTALLIN.
 DR SMART; SM00247; XTALB; 2.
 DR PROSITE; PS50915; CRYSTALLIN_BETAGAMMA; 3.
 SQ SEQUENCE 232 AA; 26781 MW; 517BDB1A29AACS8B CRC64;

Query Match 17.1%; Score 7; DB 2; Length 232;

Best Local Similarity 100.0%; Pred. No. 40;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 YKIFLPD 10

DB 41 YKIFLPD 47

Search completed: February 1, 2005, 15:31:06
 Job time : 72.5556 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:53:01 ; Search time 69.7778 Seconds
(without alignments)
210.782 Million cell updates/sec

Title: SEQ5SER1694

Sequence: 1 LKLCCKMFGAENPDPFVPL.....TAVKLIAPERKEKVLGSA 41

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 2002273 seqs, 358729299 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1316031

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : A GeneSeq_23Sep04:*

1: geneseqp1980s:*\n2: geneseqp1990s:*\n3: geneseqp2000s:*\n4: geneseqp2001s:*\n5: geneseqp2002s:*\n6: geneseqp2003as:*\n7: geneseqp2003bs:*\n8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	515	2	AAW54099 Homo sapi
2	20	48.8	1149	7	ADSO8012 Novel pro
3	20	48.8	2144	4	AAH85029 Protein e
4	7	17.1	80	4	AAW91516 Human imm
5	7	17.1	118	4	AAH87229 Human imm
6	7	17.1	308	6	ABU24428 Protein e
7	7	17.1	534	8	ADJ49159 Oil-asso
8	7	17.1	534	8	ADJ48818 Oil-asso
9	7	17.1	976	8	ADN73151 Thale cre
10	7	17.1	1110	6	ABH68350 Photornab
11	7	17.1	2000	8	ADN04625 Antipsori
12	7	17.1	4618	4	AAW39043 Human pol
13	7	17.1	4618	4	ADL46346 D-alanine
14	6	14.6	32	4	AAH89273 Human imm
15	6	14.6	42	4	AAW73154 Human bon
16	6	14.6	42	4	ABG54870 Human liv
17	6	14.6	42	5	ABG43000 Human pep
18	6	14.6	44	4	AAW15639 Peptide #
19	6	14.6	44	4	ABH34633 Peptide #
20	6	14.6	44	4	AAW28139 Peptide #
21	6	14.6	44	4	ABD29466 Peptide #
22	6	14.6	44	4	ABD20051 Protein #
23	6	14.6	44	4	AAW67825 Human bon
24	6	14.6	44	4	AAW55431 Human bra
25	6	14.6	44	4	ABG49464 Human liv

26	6	14.6	44	4	AAW03380 Peptide #
27	6	14.6	44	5	ABG37371 Human pep
28	6	14.6	47	4	AAW20415 Peptide #
29	6	14.6	47	4	ABH41137 Peptide #
30	6	14.6	47	4	AAW34913 Peptide #
31	6	14.6	47	4	ABH25176 Protein #
32	6	14.6	47	4	AAW74797 Human bon
33	6	14.6	47	4	AAW61993 Human bra
34	6	14.6	47	4	ABG56579 Human liv
35	6	14.6	47	5	ABG44588 Human pep
36	6	14.6	47	7	ADC89625 Human PMS
37	6	14.6	50	4	ABH03116 Human mus
38	6	14.6	50	6	ABU12410 Novel hum
39	6	14.6	50	8	ADJ28436 Human mus
40	6	14.6	51	3	AAW01477 Human sec
41	6	14.6	61	5	ABJ10298 Human lun
42	6	14.6	61	8	ADK47657 Streptoco
43	6	14.6	62	6	ABU20127 Protein e
44	6	14.6	63	4	AAW58663 Propionib
45	6	14.6	63	6	ABW55182 Propionib

ALIGNMENTS

RESULT 1

AAW54099 standard; protein; 515 AA.

AAW54099;

28-SEP-1998 (first entry)

Homo sapiens BAP28 sequence.

BARD1; ring protein; BRCA1; breast cancer; risk; diagnosis.

Homo sapiens.

W09812327-A2.

26-MAR-1998.

19-SEP-1997; 97WO-US016842.

20-SEP-1996; 96US-0025296P.

03-APR-1997; 97US-0042611P.

04-APR-1997; 97US-0042985P.

(TEXA) UNIV TEXAS SYSTEM.

Bowcock AM, Baer R;

WPI, 1998-230317/20.

N-PSDB; AAW24135.

DNA sequence encoding BARD1, B123, BE2, BE14, BE31 or BE445 - which as breast cancer antigen, BRCA1, binding proteins are useful to identify patient having or at risk of developing cancer.

Disclosure, Page 287-288; 348pp; English.

The sequence is that of a protein which can be used in the preparation of the recombinant breast cancer antigen, BRCA1, binding proteins BARD1, B123, BE2, BE14, BE31 or BE445, or a composition for the detection of a BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid sequence, specifically a wild type BARD1 composition for the detection or purification of BRCA1, useful to identify a patient having, or at risk of developing cancer. BARD1 can be used in the preparation of an anti-BARD1 antibody, and in the detection and purification of a BRCA1 protein. BARD1, B123, BE2, BE14, BE31 or BE445 can be used in the identification of a binding protein against or antagonist that alters the binding of BARD1, B123, BE2, BE14, BE31 or BE445 to BRCA1 or the biological activity of the BRCA1-BARD1,

CC B123, B23, BE14, BE31 or BE445 complex. The antibodies can be used to
 CC detect BARD1, B123, B23, BE14, BE31 or BE445, a specific anti-BARD1
 CC antibody can be used to identify a patient having or at risk of
 CC developing cancer
 CC
 SQ Sequence 515 AA;

Query Match 100.0%; Score 41; DB 2; Length 515;
 Best Local Similarity 100.0%; Pred. No. 6.4e-34;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DY 1 LKLLCKNFGAENPDPPVPLSTAVKLIAPRKEKKNVLGSA 41
 45 LKLLCKNFGAENPDPPVPLSTAVKLIAPRKEKKNVLGSA 85

RESULT 2

ID ADE08012 standard; protein; 1149 AA.

AC ADE08012;

DT 29-JAN-2004 (first entry)

DE Novel protein (useful for identifying genetic disorders) #167.

KW novel gene; novel protein; tissue marker; molecular weight marker;
 KM chromosome marker; genetic disorder.

OS Unidentified.

PN WO2003054152-A2.

PD 03-JUL-2003.

PF 10-DEC-2002; 2002WO-US039555.

PR 10-DEC-2001; 2001US-0339739P.

PR 11-DEC-2001; 2001US-0339453P.

PR 14-MAR-2002; 2002US-0365091P.

PR 12-APR-2002; 2002US-0372381P.

PR 22-APR-2002; 2002US-0012855P.

PR 24-APR-2002; 2002US-0376045P.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
 PI Ghosh M, Xue AJ, Wehrman T, Meng G, Zhou P, Dmanac RT, Wang Z;
 PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ,

DR WPI; 2003-569235/53.

DR N-PSDB; ADE07101.

PT New polynucleotides, useful for expressing recombinant proteins for
 PT analysis, characterization or therapeutic use, or as markers for tissues
 PT in which the corresponding protein is preferentially expressed.

PS Claim 20; SEQ ID NO 1078; 1177bp; English.

CC The invention comprises the amino acid and coding sequences of novel
 CC proteins. The DNA and protein sequences of the invention are useful as:
 CC markers for tissues in which the corresponding protein is preferentially
 CC expressed; as molecular weight markers on gels; as chromosome markers or
 CC tags; to identify chromosomes or to map related gene positions; and to
 CC compare with endogenous DNA sequences in patients to identify potential
 CC genetic disorders. The present amino acid sequence represents a protein
 CC of the invention.

SQ Sequence 1149 AA;

Query Match 48.8%; Score 20; DB 7; Length 1149;

Best Local Similarity 100.0%; Pred. No. 6.6e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DY 1 LKLLCKNFGAENPDPPVPL 20
 732 LKLLCKNFGAENPDPPVPL 751

RESULT 3

ID AAB85029 standard; protein; 2144 AA.

AC AAB85029;

DT 06-AUG-2001 (first entry)

DE Protein encoded by BAP28 cDNA consisting of exons 1 to 45.

KW BAP28; prostate; tumour; cancer; diagnostic; genetic analysis.

OS Homo sapiens.

PH Key Location/Qualifiers

FT Misc-difference 1694 /label= Ser or Asn

FT Misc-difference 1854 /label= Ala or Val

FT Misc-difference 1967 /label= Asp or Asn

FT Misc-difference 2017 /label= Gly or Glu

PN WO200100669-A2.

PD 04-JAN-2001.

PF 23-JUN-2000; 2000WO-IB001183.

PR 25-JUN-1999; 99US-0141323P.

PR 18-JAN-2000; 2000US-0176880P.

PA (GERT) GENSET.

PI Barry C, Bougueleret L, Chumakov I, Cohen-Akenine A;

DR N-PSDB; AAF83909, AAF83910.

DR WPI; 2001-367032/38.

PS Claim 14; Page 297-304; 349pp; English.

CC The invention is directed to BAP28 polypeptides, BAP28 polynucleotide
 CC sequences and regulatory region located at the 3' and 5' ends of the
 CC BAP28 coding region. The BAP28 polypeptides can be expressed by standard
 CC recombinant methodology. BAP28 polynucleotides and polypeptides have been
 CC found to be over expressed in prostate tumour cells, therefore levels of
 CC BAP28 expression and/or activity may be assayed (e.g. by polymerase chain
 CC reaction (PCR)) to diagnose patient suffering from or susceptible to
 CC prostate cancer. Antibodies specific for the BAP28 polypeptides are
 CC useful as diagnostic reagents. Biallelic markers of the BAP28 gene are
 CC useful in genetic analysis. The present sequence represents a protein
 CC encoded by a first cDNA sequence of the BAP28 gene consisting of the
 CC exons 1 to 45.

SQ Sequence 2144 AA;

Query Match 48.8%; Score 20; DB 4; Length 2144;
 Best Local Similarity 100.0%; Pred. No. 1.1e-11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251888P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0251719P.
PR 08-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251899P.
PR 08-DEC-2000; 2000US-0251909P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX
XX MPI; 2001-483426/52.
DR N-PDB; AAK64297.
XX
XX Nucleic acids encoding human immune/haematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Claim 11; SEQ ID NO 19109; 3071BP + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM62170 to AAM91921. (I) Have cytostratic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC protein and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK4703
CC to AAK67694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM62169
CC represent sequences used in the exemplification of the present invention
XX
XX SQ Sequence 80 AA;
XX
XX Query Match 17.1%; Score 7; DB 4; Length 80;
XX Best Local Similarity 100.0%; Fred. No. 24;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 18 PVLSTAV 24
XX |||||
XX 59 PVLSTAV 65
XX
XX DB
XX
XX RESULT 5
XX AAM87229
XX ID AAM87229 standard; protein; 118 AA.
XX
XX AC AAM87229;
XX
XX DT 07-NOV-2001 (first entry)
XX
XX DE Human immune/haematopoietic antigen SEQ ID NO:14822.
XX
XX DE Human immune/haematopoietic antigen SEQ ID NO:14822.
XX
XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostratic; gene therapy; vaccine; metastasis.
XX
XX OS Homo sapiens.
XX
XX

PN WO200157182-A2.
XX
XX PD 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001354.
XX
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226686P.
XX 22-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.
XX 08-SEP-2000; 2000US-0231242P.
XX 08-SEP-2000; 2000US-0231243P.
XX 08-SEP-2000; 2000US-0231244P.
XX 08-SEP-2000; 2000US-0231413P.
XX 08-SEP-2000; 2000US-0231414P.
XX 08-SEP-2000; 2000US-0231414P.
XX 08-SEP-2000; 2000US-0232080P.
XX 08-SEP-2000; 2000US-0232081P.
XX 12-SEP-2000; 2000US-0231968P.
XX 14-SEP-2000; 2000US-0232397P.
XX 14-SEP-2000; 2000US-0232398P.
XX 14-SEP-2000; 2000US-0232399P.
XX 14-SEP-2000; 2000US-0232400P.
XX 14-SEP-2000; 2000US-0232401P.
XX 14-SEP-2000; 2000US-0233063P.
XX 14-SEP-2000; 2000US-0233064P.
XX 14-SEP-2000; 2000US-0233065P.
XX 21-SEP-2000; 2000US-0234223P.
XX 21-SEP-2000; 2000US-0234274P.
XX 25-SEP-2000; 2000US-0234997P.
XX 25-SEP-2000; 2000US-0234998P.
XX 26-SEP-2000; 2000US-0235484P.
XX 27-SEP-2000; 2000US-0235434P.
XX 27-SEP-2000; 2000US-0235636P.

PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 02-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM,
 XX

XX WPI, 2001-483426/52.
 DR N-PSDB; AAK60010.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides.
 PT Useful for preventing, diagnosing and/or treating cancers and metastasis.
 XX
 PS Claim 11; SEQ ID NO 14822; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAK62170 to AAK91921. (I) have cytosolic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK67694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169
 CC represent sequences used in the exemplification of the present invention
 XX
 SQ Sequence 118 AA;
 Query Match 17.1%; Score 7; DB 4; Length 118;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 31 RKEKNV 37
 Db 52 RKEKNV 58
 RESULT 6
 ABU24428
 ID ABU24428 standard; protein, 308 AA.
 AC ABU24428;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #9955.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Clostridium botulinum.
 XX
 FN WO20027183-A2.
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;
 XX
 DR WPI, 2003-029926/02.
 DR N-PSDB; ACA28298.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.

PS Claim 25; SEQ ID NO 52352; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway of
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

SO Sequence 308 AA;

Query Match 17.1%; Score 7; DB 6; Length 308;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 KNVLGS 41
 DB 298 KNVLGS 304

RESULT 7
 ADJ49159
 ID ADJ49159 standard; protein; 534 AA.

AC ADJ49159;

DT 06-MAY-2004 (first entry)

DE Oil-associated gene related protein #659.

KW oil-associated gene; transgenic; enhanced seed oil; vegetable oil.

OS Unidentified.

PN US2004025202-A1.

PD 05-FEB-2004.

PF 14-MAR-2003; 2003US-00389566.

PR 15-MAR-2002; 2002US-0365301P.

PR 26-JUN-2002; 2002US-0391786P.

PR 26-JUN-2002; 2002US-0392018P.

PA (LAURIE C C.

PA (RAVA/) RAVANELLO M.
 PA (SAVA/) SAVAGE T.
 PA (LEDE/) LEDEAUX J R.
 PA (ROGE/) ROGERS J A.

PI Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;
 DR WPI; 2004-142683/14.

PT Novel recombinant DNA construct comprising a promoter functional in
 PT plants operably linked to an oil-associated gene for producing transgenic
 PT plant seed.

PS Example 3; SEQ ID NO 1163; 22pp; English.

CC The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in plants operably linked to an oil-associated gene.
 CC The construct is useful for transgenic plant seed which has in its genome
 CC the construct, that is functional in the plant to transcribe the oil-
 CC associated gene. The transgenic plant seed grows into a plant having
 CC enhanced seed oil as compared to wild type. The construct is useful for
 CC producing hybrid maize seed. The transgenic plant seed is useful for
 CC producing vegetable oil. The present sequence represents the amino acid
 CC sequence of an oil-associated gene related protein.

SO Sequence 534 AA;

Query Match 17.1%; Score 7; DB 8; Length 534;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 LSTAVKL 26
 DB 303 LSTAVKL 309

RESULT 8
 ADJ48818
 ID ADJ48818 standard; protein; 534 AA.

AC ADJ48818;

DT 06-MAY-2004 (first entry)

DE Oil-associated gene related protein #318.

KW oil-associated gene; transgenic; enhanced seed oil; vegetable oil.

OS Unidentified.

PN US2004025202-A1.

PD 05-FEB-2004.

PF 14-MAR-2003; 2003US-00389566.

PR 15-MAR-2002; 2002US-0365301P.

PR 26-JUN-2002; 2002US-0391786P.

PR 26-JUN-2002; 2002US-0392018P.

PA (LAURIE C C.
 PA (RAVA/) RAVANELLO M.
 PA (SAVA/) SAVAGE T.
 PA (LEDE/) LEDEAUX J R.
 PA (ROGE/) ROGERS J A.

PI Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;

DR WPI; 2004-142683/14.

PT Novel recombinant DNA construct comprising a promoter functional in
 PT plants operably linked to an oil-associated gene for producing transgenic
 PT plant seed.

XX Example 3; SEQ ID NO 822; 22pp; English.
 PS
 CC The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in plants operably linked to an oil-associated gene.
 CC The construct is useful for transgenic plant seed which has in its genome
 CC the construct, that is functional in the plant to transcribe the oil-
 CC associated gene. The transgenic plant seed grows into a plant having
 CC enhanced seed oil as compared to wild type. The construct is useful for
 CC producing hybrid maize seed. The transgenic plant seed is useful for
 CC producing vegetable oil. The present sequence represents the amino acid
 CC sequence of an oil-associated gene related protein.
 CC
 SQ Sequence 534 AA;
 Query Match 17.1%; Score 7; DB 8; Length 534;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 20 LSTAVKL 26
 DB 303 LSTAVKL 309
 RESULT 9
 ID ADN73151
 AC ADN73151 standard; protein; 976 AA.
 XX
 AC ADN73151;
 XX
 DT 15-JUL-2004 (first entry)
 XX
 DE Thale cress protein upregulated in E2Fa/Dpa expressing plants SeqID 1046.
 XX
 KW plant; transgenic; E2Fa/Dpa transcription factor; growth regulator;
 KW animal feed product; thale cress; cell wall biosynthesis;
 KW nitrogen metabolism; carbon metabolism.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO2004035798-A2.
 XX
 PD 29-APR-2004.
 XX
 PF 20-OCT-2003; 2003WO-EP011658.
 XX
 PR 18-OCT-2002; 2002EP-00079408.
 XX
 PA (CROP-) CROPDESIGN NV.
 XX
 PI Inze D, De Veylder L, Vlieghe K;
 XX
 DR WPI; 2004-348466/32.
 DR N-PSDB; ADN73150.
 XX
 PT Altering plant characteristics, useful for producing plants for enzyme or
 PT pharmaceutical production comprises modifying in a plant, expression of
 PT one or more nucleic acids and/or modifying level or activity of one or
 PT more proteins.
 XX
 PS Claim 1; SEQ ID NO 1046; 134pp; English.
 XX
 CC This invention relates to a novel method for altering one or more plant
 CC characteristics. Specifically, it refers to identifying genes that are up
 CC - or down-regulated in transgenic plants overexpressing the heterodimeric
 CC E2Fa/Dpa transcription factor of Arabidopsis and using these sequences to
 CC alter plant characteristics accordingly. The present invention describes
 CC generating transgenic plants for the production of growth regulators,
 CC enzymes, therapeutics, pharmaceuticals and animal feed products, where
 CC the altered plant characteristics are selected from increased yield or
 CC biomass, enhanced survival capacity, stress tolerance, plant architecture
 CC or physiology, altered endoreplication, biochemistry, signal
 CC transduction, storage lipid mobilisation and/or altered photosynthesis,

CC each relative to the corresponding wild type plants. Accordingly, these
 CC sequences can also be useful as positive or negative selectable markers
 CC during transformation of cells or tissues. The identified genes play a
 CC role in a variety of biological processes such as DNA replication, cell
 CC wall biosynthesis, nitrogen and/or carbon metabolism or they function as
 CC transcription factors. This polypeptide sequence is thale cress protein
 CC expressed by a gene upregulated 1.3 fold or more in plants overexpressing
 CC the E2Fa/Dpa transcription factor, given in an exemplification of the
 CC invention.
 CC
 SQ Sequence 976 AA;
 Query Match 17.1%; Score 7; DB 8; Length 976;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 32 KEKNVL 38
 DB 474 KEKNVL 480
 RESULT 10
 ID ABM68350
 AC ABM68350 standard; protein; 1110 AA.
 XX
 AC ABM68350;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Photorhabdus luminescens protein sequence #1447.
 XX
 KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 KW whooping cough.
 XX
 OS Photorhabdus luminescens.
 XX
 PN WO200294867-A2.
 XX
 PD 28-NOV-2002.
 XX
 PF 07-FEB-2002; 2002WO-IB003040.
 XX
 PR 07-FEB-2001; 2001FR-00001659.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A,
 PI Buchrieser C;
 XX
 DR WPI; 2003-148459/14.
 XX
 PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
 XX
 PS Claim 2; SEQ ID NO 1447; 1205pp; French.
 XX
 CC The invention relates to the isolation of genes and their encoded
 CC proteins from Photorhabdus luminescens. The isolated sequences are
 CC sources of probes and primers for detecting the genome of P. luminescens
 CC and related species; to study polymorphisms; for gene analysis and for
 CC detection/amplification of the genes. Antibodies (Ab) raised against the
 CC polypeptides encoded by the genes are used for detection/identification
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
 CC carry a gene-containing vector are used to select compounds that
 CC modulate, regulate, induce or inhibit expression of the genes in plants,
 CC animals or microorganisms other than P. luminescens and are able to alter
 CC response or sensitivity to toxins and antibiotics produced by P.
 CC luminescens. Cells transformed to express the genes are useful for
 CC recombinant production of the proteins, particularly toxins and
 CC antibacterials useful as insecticides, bactericides and fungicides. The

CC genes, proteins, vectors containing the genes and Ab are also useful
 CC therapeutically (to treat microbial infection by bacteria or fungi that
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
 CC biopesticides. Other uses of the genes and the proteins are as virulence
 CC factors and for identifying targets of human diseases for which P.
 CC luminescens is a model (particularly plague and whooping cough). This
 CC sequence represents one of the isolated P. luminescens proteins
 XX
 SQ Sequence 1110 AA;

Query Match 17.1%; Score 7; DB 6; Length 1110;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 FVPVLSST 22
 |||||
 Db 784 FVPVLSST 790

RESULT 11

ADN04625
 ID ADN04625 standard; protein; 2000 AA.

XX
 AC ADN04625;

XX 01-JUL-2004 (first entry)

XX Antipsoriatic protein sequence #498.

XX Antipsoriatic; gene therapy; psoriasis; diagnosis.

XX Homo sapiens.

XX WO2004028479-A2.

XX 08-APR-2004.

XX 25-SEP-2003; 2003WO-US030907.

XX 25-SEP-2002; 2002US-0414006P.

XX (GENTH) GENENTECH INC.

XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI,
 PI Wu JD;

XX WPI; 2004-305105/28.

XX N-PSDB; ADN04624.

XX New PRO nucleic acid or polypeptide, useful for preparing a
 PT pharmaceutical composition for diagnosing or treating psoriasis in a
 PT mammal.

XX Claim 9; SEQ ID NO 1019; 3069pp; English.

XX The invention relates to novel polynucleotide and polypeptides for
 CC treating psoriasis or a sequence having at least 80% identity to the
 CC above sequences. The nucleic acid is useful for preparing a composition
 CC for diagnosing or treating psoriasis in a mammal. This sequence
 CC corresponds to one of the polypeptides of the invention.

XX Sequence 2000 AA;

XX Query Match 17.1%; Score 7; DB 8; Length 2000;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 APERKEE 34
 |||||
 Db 1333 APERKEE 1339

XX Sequence 2000 AA;

XX Query Match 17.1%; Score 7; DB 8; Length 2000;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 APERKEE 34
 |||||
 Db 1333 APERKEE 1339

XX Sequence 2000 AA;

QY 28 APERKEE 34
 |||||
 Db 1333 APERKEE 1339

RESULT 12

AAAM39043
 ID AAAM39043 standard; protein; 4618 AA.

XX
 AC AAAM39043;

XX 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 2188.

XX Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US034263.

XX 23-DEC-1999; 99US-00471275.

XX 21-JAN-2000; 2000US-00488725.

XX 25-APR-2000; 2000US-00552317.

XX 20-JUN-2000; 2000US-00598042.

XX 19-JUL-2000; 2000US-00620312.

XX 03-AUG-2000; 2000US-00653450.

XX 14-SEP-2000; 2000US-00662191.

XX 19-OCT-2000; 2000US-00693036.

XX 29-NOV-2000; 2000US-00727344.

XX (HYSE-) HXSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao Q,
 PI Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX N-PSDB; AA158199.

XX Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.

XX Example 4; SEQ ID NO 2188; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and the
 CC encoded polypeptides (AAAM38642-AAAM42213) with noctropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification

XX Sequence 4618 AA;

XX Query Match 17.1%; Score 7; DB 4; Length 4618;
 Best Local Similarity 100.0%; Pred. No. 5.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 APERKEE 34
 |||||
 Db 1544 APERKEE 1550

XX Sequence 4618 AA;

QY 28 APERKEE 34
 |||||
 Db 1544 APERKEE 1550

XX Sequence 4618 AA;

QY 28 APERKEE 34
 |||||
 Db 1544 APERKEE 1550

XX Sequence 4618 AA;

QY 28 APERKEE 34
 |||||
 Db 1544 APERKEE 1550

XX Sequence 4618 AA;

QY 28 APERKEE 34
 |||||
 Db 1544 APERKEE 1550

RESULT 13
ADL46346
ID ADL46346 standard; peptide, 21 AA.
XX
AC ADL46346;
DT 20-MAY-2004 (first entry)
XX
DE D-alanine:D-alanine adding enzyme epitope #1.
XX
KM epitope; antibacterial;
KM UDP-N-acetylglucosamine 1-carboxyvinyl transferase-1;
KM CTP:UMP-3-deoxy-D-manno-ocutulosonate transferase;
KM UDP-N-acetylmuramylalanyl-D-glutamate-2-6-diaminopimelate ligase;
KM D-alanine-D-alanine adding enzyme; D-alanine-D-alanine ligase;
KM UDP-N-acetylpyruvoylglucosamine reductase;
KM UDP-N-acetylglucosamine pyrophosphorylase;
KM UDP-N-acetylmuramoylalanyl-D-glutamate ligase;
KM DP-N-acetylmuramate:alanine ligase; aspartate semialdehyde dehydrogenase;
KM UDP-N-acetylmuramoylalanyl-D-glutamate; X-ray diffraction analysis;
enzyme.
XX
XX Pseudomonas aeruginosa.
XX
XX WO2003087353-A2.
XX
PD 23-OCT-2003.
XX
XX 08-APR-2003; 2003WO-CA000481.
XX
PF 08-APR-2002; 2002US-0370899P.
XX
PR 08-APR-2002; 2002US-0370915P.
XX
PR 09-APR-2002; 2002US-0371107P.
XX
PR 09-APR-2002; 2002US-0371185P.
XX
PR 31-MAY-2002; 2002US-0385426P.
XX
PR 06-JUN-2002; 2002US-0386283P.
XX
PR 01-AUG-2002; 2002US-0400348P.
XX
PR 06-NOV-2002; 2002US-0424395P.
XX
PR 08-NOV-2002; 2002US-0425200P.
XX
PR 24-DEC-2002; 2002US-0436345P.
XX
PR 24-DEC-2002; 2002US-0436349P.
XX
PR 26-DEC-2002; 2002US-0436568P.
XX
PR 27-DEC-2002; 2002US-043675P.
XX
PR 27-DEC-2002; 2002US-0436734P.
XX
PR 27-DEC-2002; 2002US-0436885P.
XX
PR 27-DEC-2002; 2002US-0436893P.
XX
PR 27-DEC-2002; 2002US-0436893P.
XX
PR 27-DEC-2002; 2002US-0436900P.
XX
PR 30-DEC-2002; 2002US-0437013P.
XX
XX (AFPI-) AFFINIUM PHARM INC.
XX
PI Edwards A, Dharamsi A, Vedadi M, Domagala M, Houston S, Awrey D;
PI Beattie B, Mansoury K, Ouyang H, Vallee F, Richards D, Nethery K;
PI Vitag C, Buzadzija K, Pinder B, Alam WZ, Tai M, Canadian V;
PI Kanagareja D, Thalakkada R;
XX
XX MPI; 2003-865361/80.
XX
DR
XX
PT New recombinant bacterial enzymes involved in cell membrane biogenesis,
PT useful for designing potential antibacterial agents.
XX
PS Disclosure; SEQ ID NO 64; 407pp; English.
XX
XX The invention relates to isolated, recombinant polypeptides (I) that have
XX at least one activity of specified bacterial enzymes involved in cell
XX membrane biogenesis. (I) are: UDP-N-acetylglucosamine 1-carboxyvinyl
XX transferase-1 of Streptococcus pneumoniae (S.a.); Pseudomonas aeruginosa
XX (P.a.) or Staphylococcus aureus (S.a.); CTP:UMP-3-deoxy-D-manno-
XX octulosonate transferase of Escherichia coli (E.c.) or Haemophilus
XX influenzae (H.i.); UDP-N-acetylmuramylalanyl-D-glutamate-2,6-

CC diaminopimelate ligase of P.a.; D-alanine:D-alanine adding enzyme of S.a.
CC or P.a.; D-alanine-D-alanine ligase of Enterococcus faecalis (E.f.); UDP-N-
CC -acetylpyruvoylglucosamine reductase of P.a. or H.i.; UDP-N-
CC acetylglucosamine pyrophosphorylase of E.f., H.i. or S.a.; DP-N-
CC acetylmuramoylalanyl-D-glutamate ligase of E.f. or H.i.; DP-N-
CC acetylmuramate:alanine ligase of E.c.; and aspartate semialdehyde
CC dehydrogenase of H.i. and UDP-N-acetylmuramoylalanyl-D-glutamate (sic) of
CC H.i. Crystalline (I) are used to determine (by X-ray diffraction
CC analysis) the structural coordinates of (I), and these then used to
CC design modulators of (I), potential therapeutic agents for treating
CC diseases caused by the specified bacteria. This sequence represents an
CC epitope from one of the proteins of the invention.
XX
XX SQ Sequence 21 AA;
XX
XX Query Match 14.6%; Score 6; DB 7; Length 21;
XX Best Local Similarity 100.0%; Pred. No. 89;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 15 PFVPTL 20
DB 14 PFVPTL 19
RESULT 14
AAM89273
ID AAM89273 standard; protein, 32 AA.
XX
XX AAM89273;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen SEQ ID NO:16866.
XX
KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KM cytostatic; gene therapy; vaccine; metabolite.
XX
OS Homo sapiens.
XX
XX WO200157182-A2.
XX
PN 09-AUG-2001.
XX
PD 17-JAN-2001; 2001WO-US001354.
XX
PF 31-JAN-2000; 2000US-0179065P.
XX
PR 04-FEB-2000; 2000US-0180628P.
XX
PR 24-FEB-2000; 2000US-0184664P.
XX
PR 02-MAR-2000; 2000US-0186350P.
XX
PR 16-MAR-2000; 2000US-0189874P.
XX
PR 17-MAR-2000; 2000US-0190076P.
XX
PR 18-APR-2000; 2000US-0198123P.
XX
PR 19-MAY-2000; 2000US-0205515P.
XX
PR 07-JUN-2000; 2000US-0209467P.
XX
PR 28-JUN-2000; 2000US-0214886P.
XX
PR 30-JUN-2000; 2000US-0215135P.
XX
PR 07-JUL-2000; 2000US-0216647P.
XX
PR 07-JUL-2000; 2000US-0216880P.
XX
PR 11-JUL-2000; 2000US-0217487P.
XX
PR 11-JUL-2000; 2000US-0217496P.
XX
PR 14-JUL-2000; 2000US-0218290P.
XX
PR 26-JUL-2000; 2000US-0220963P.
XX
PR 26-JUL-2000; 2000US-0220964P.
XX
PR 14-AUG-2000; 2000US-0224518P.
XX
PR 14-AUG-2000; 2000US-0224519P.
XX
PR 14-AUG-2000; 2000US-0225213P.
XX
PR 14-AUG-2000; 2000US-0225214P.
XX
PR 14-AUG-2000; 2000US-0225266P.
XX
PR 14-AUG-2000; 2000US-0225267P.
XX
PR 14-AUG-2000; 2000US-0225268P.
XX
PR 14-AUG-2000; 2000US-0225270P.
XX
PR 14-AUG-2000; 2000US-0225447P.
XX
PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226779P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0226924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0232403P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0234984P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.

PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251866P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
PI WPI, 2001-483426/52.
XX N-PSDB; AAK62054.
DR
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Claim 11; SEQ ID NO 16866; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 32 AA;

Query Match 14.6%; Score 6; DB 4; Length 32;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 35 KNTVGS 40

Db 12 KNTLGS 17

RESULT 15

AAM73154
ID AAM73154 standard; protein; 42 AA.

XX AAM73154;

DT 06-NOV-2001 (first entry)

DE Human bone marrow expressed probe encoded protein SEQ ID NO: 33460.

XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukemia; lymphoma; myeloma.

XX Homo sapiens.

PN WO200157276-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000668.

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632356.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

PI MPI; 2001-488900/53.

DR Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human bone marrow.

XX Example 4; SEQ ID NO 33460; 658bp + Sequence Listing; English.

PS The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human

CC bone marrow. They can be used to measure gene expression in bone marrow

CC samples, which may enable the improved diagnosis and treatment of cancers

CC such as lymphoma, leukemia and myeloma. The present sequence is a

CC protein encoded by one of the probes of the invention

XX Sequence 42 AA;

SO Query Match 14.6%; Score 6; DB 4; Length 42;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 VPVLST 22

Db 16 VPVLST 21

Search completed: February 1, 2005, 15:20:08
Job time : 73.7778 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:58:28 ; Search time 17.5556 Seconds
(without alignments)
154.882 Million cell updates/sec

Title: SEQSERVER1694

Perfect score: 41
Sequence: 1 LKLLCKNFGAENPDPPFVVL.....TAVKLIAPERKEKVLGSA 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size: 0

Total number of hits satisfying chosen parameters: 290433

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Issued Patents AA:*

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTOS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	14.6	51	4	US-09-513-999C-5558
2	6	14.6	61	4	US-09-583-110-4172
3	6	14.6	81	4	US-09-621-976-6450
4	6	14.6	87	4	US-09-513-999C-6616
5	6	14.6	88	4	US-09-248-796A-25228
6	6	14.6	90	4	US-09-252-991A-21131
7	6	14.6	99	4	US-09-270-767-35660
8	6	14.6	99	4	US-09-270-767-50877
9	6	14.6	127	4	US-09-270-767-38390
10	6	14.6	127	4	US-09-270-767-53607
11	6	14.6	129	6	5229115-2
12	6	14.6	133	4	US-09-708-200-17
13	6	14.6	135	4	US-09-248-796A-22195
14	6	14.6	147	4	US-09-710-279-1214
15	6	14.6	154	1	US-08-446-908-2
16	6	14.6	154	1	US-08-231-205A-2
17	6	14.6	154	2	US-08-871-161-2
18	6	14.6	169	4	US-09-248-796A-19969
19	6	14.6	176	4	US-09-489-039A-11200
20	6	14.6	205	4	US-09-252-991A-26704
21	6	14.6	218	4	US-09-710-279-450
22	6	14.6	218	4	US-09-248-796A-18711
23	6	14.6	237	4	US-09-248-796A-18839
24	6	14.6	253	2	US-08-685-992-10
25	6	14.6	253	2	US-09-144-925-10
26	6	14.6	264	4	US-09-788-657-24
27	6	14.6	264	4	US-09-788-657-25

28	6	14.6	269	4	US-09-252-991A-25341	Sequence 25341, A
29	6	14.6	294	4	US-09-540-236-3779	Sequence 3779, Ap
30	6	14.6	298	4	US-09-134-000C-5845	Sequence 5845, Ap
31	6	14.6	305	3	US-09-105-390-36	Sequence 36, Appl
32	6	14.6	309	4	US-09-489-039A-8203	Sequence 8203, Ap
33	6	14.6	316	4	US-09-252-991A-25345	Sequence 25345, Ap
34	6	14.6	325	3	US-09-134-001C-3551	Sequence 3551, Ap
35	6	14.6	329	4	US-09-149-476-483	Sequence 483, App
36	6	14.6	336	3	US-09-105-390-52	Sequence 52, Appl
37	6	14.6	339	4	US-09-107-532A-5514	Sequence 5514, Ap
38	6	14.6	344	4	US-09-248-796A-16383	Sequence 16383, A
39	6	14.6	347	4	US-09-252-991A-31647	Sequence 31647, A
40	6	14.6	348	1	US-08-454-196-8	Sequence 8, Appl1
41	6	14.6	348	1	US-08-454-196-17	Sequence 17, Appl
42	6	14.6	348	3	US-09-064-033-8	Sequence 8, Appl1
43	6	14.6	348	3	US-09-064-033-17	Sequence 17, Appl
44	6	14.6	348	4	US-09-291-046-8	Sequence 8, Appl1
45	6	14.6	348	4	US-09-291-046-17	Sequence 17, Appl

ALIGNMENTS

```
RESULT 1
US-09-513-999C-5558
; Sequence 5558, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Mline Edwards, J.B.
; APPLICANT: Duclet, A.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59 US2 REG
; CURRENT APPLICATION NUMBER: US/09/513, 999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 5558
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-513-999C-5558

Query Match      14.6%; Score 6; DB 4; Length 51;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      33 EKNVL 38
      |||||
Db      25 EKNVL 30

RESULT 2
US-09-583-110-4172
; Sequence 4172, Application US/09583110
; Patent No. 669703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PAT00-67A
; CURRENT APPLICATION NUMBER: US/09/583, 110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
```

SEQ ID NO 4172
 LENGTH: 61
 TYPE: PRT
 ORGANISM: Streptococcus pneumoniae
 US-09-583-110-4172

Query Match 14.6%; Score 6; DB 4; Length 61;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LCKNF 8
 Db 11 LCKNF 16

RESULT 3
 US-09-621-976-6450
 Sequence 6450, Application US/09621976
 Patent No. 6639063
 GENERAL INFORMATION:
 APPLICANT: Dumas Milne Edwards, J.B.
 APPLICANT: Jobert, S.
 APPLICANT: Giordano, J.Y.
 TITLE OF INVENTION: ESTs and Encoded Human Proteins.
 FILE REFERENCE: GENSET.054PR2
 CURRENT APPLICATION NUMBER: US/09/621,976
 CURRENT FILING DATE: 2000-07-21
 NUMBER OF SEQ ID NOS: 19335
 SOFTWARE: Patent.pm
 SEQ ID NO 6450
 LENGTH: 81
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-621-976-6450

Query Match 14.6%; Score 6; DB 4; Length 81;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 ERKEK 35
 Db 51 ERKEK 56

RESULT 4
 US-09-513-999C-4616
 Sequence 4616, Application US/09513999C
 Patent No. 6783961
 GENERAL INFORMATION:
 APPLICANT: Dumas Milne Edwards, J.B.
 APPLICANT: Duclercq, A.
 APPLICANT: Giordano, J.Y.
 TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
 Patent No. 6783961
 FILE REFERENCE: 59, US2, REG
 CURRENT APPLICATION NUMBER: US/09/513,999C
 CURRENT FILING DATE: 2000-02-24
 PRIOR APPLICATION NUMBER: US 60/122,487
 PRIOR FILING DATE: 1999-02-26
 NUMBER OF SEQ ID NOS: 36681
 SOFTWARE: Patent.pm
 SEQ ID NO 4616
 LENGTH: 87
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: SIGNAL
 LOCATION: -48...-1
 OTHER INFORMATION: score 3.8
 OTHER INFORMATION: seq NSLLLLCLLYYP/HS
 FEATURE:
 NAME/KEY: UNSURE
 LOCATION: 7

OTHER INFORMATION: Xaa=Asp or Glu
 FEATURE:
 NAME/KEY: UNSURE
 LOCATION: 8
 OTHER INFORMATION: Xaa= * or Cys or Phe or Leu or Trp or Tyr
 FEATURE:
 NAME/KEY: UNSURE
 LOCATION: 21
 OTHER INFORMATION: Xaa=Ala or Asp or Glu or Gly or Ile or Lys or Met or Asn or Arg o
 FEATURE:
 NAME/KEY: UNSURE
 LOCATION: 22
 OTHER INFORMATION: Xaa=Cys or Phe
 US-09-513-999C-4616

Query Match 14.6%; Score 6; DB 4; Length 87;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LCKNF 8
 Db 15 LCKNF 20

RESULT 5
 US-09-248-796A-25228
 Sequence 25228, Application US/09248796A
 Patent No. 6747137
 GENERAL INFORMATION:
 APPLICANT: Keith Weinstock et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
 FILE REFERENCE: 107196.132
 CURRENT APPLICATION NUMBER: US/09/248,796A
 CURRENT FILING DATE: 1999-02-12
 PRIOR APPLICATION NUMBER: US 60/074,725
 PRIOR FILING DATE: 1998-02-13
 PRIOR APPLICATION NUMBER: US 60/096,409
 PRIOR FILING DATE: 1998-08-13
 NUMBER OF SEQ ID NOS: 28208
 SEQ ID NO 25228
 LENGTH: 88
 TYPE: PRT
 ORGANISM: Candida albicans
 US-09-248-796A-25228

Query Match 14.6%; Score 6; DB 4; Length 88;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 ERKEK 35
 Db 35 ERKEK 40

RESULT 6
 US-09-252-991A-21131
 Sequence 21131, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 21131
 LENGTH: 90

```
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21131

Query Match      14.6%; Score 6; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      20 LSTAVK 25
Db      63 LSTAVK 68

RESULT 7
US-09-270-767-35660
/ Sequence 35660, Application US/09270767
/ Patent No. 6703491
/ GENERAL INFORMATION:
/ APPLICANT: Homburger et al.
/ TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
/ FILE REFERENCE: File Reference: 7326-094
/ CURRENT APPLICATION NUMBER: US/09/270,767
/ CURRENT FILING DATE: 1999-03-17
/ NUMBER OF SEQ ID NOS: 62517
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 35660
/ LENGTH: 99
/ TYPE: PRT
/ ORGANISM: Drosophila melanogaster
/ FEATURE:
/ OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-35660

Query Match      14.6%; Score 6; DB 4; Length 99;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      13 PDPFVP 18
Db      86 PDPFVP 91

RESULT 8
US-09-270-767-50877
/ Sequence 50877, Application US/09270767
/ Patent No. 6703491
/ GENERAL INFORMATION:
/ APPLICANT: Homburger et al.
/ TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
/ FILE REFERENCE: File Reference: 7326-094
/ CURRENT APPLICATION NUMBER: US/09/270,767
/ CURRENT FILING DATE: 1999-03-17
/ NUMBER OF SEQ ID NOS: 62517
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 50877
/ LENGTH: 99
/ TYPE: PRT
/ ORGANISM: Drosophila melanogaster
/ FEATURE:
/ OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-50877

Query Match      14.6%; Score 6; DB 4; Length 99;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      13 PDPFVP 18
Db      86 PDPFVP 91

RESULT 9
US-09-270-767-38390

/ Sequence 38390, Application US/09270767
/ Patent No. 6703491
/ GENERAL INFORMATION:
/ APPLICANT: Homburger et al.
/ TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
/ FILE REFERENCE: File Reference: 7326-094
/ CURRENT APPLICATION NUMBER: US/09/270,767
/ CURRENT FILING DATE: 1999-03-17
/ NUMBER OF SEQ ID NOS: 62517
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 38390
/ LENGTH: 127
/ TYPE: PRT
/ ORGANISM: Drosophila melanogaster
/ FEATURE:
/ OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-38390

Query Match      14.6%; Score 6; DB 4; Length 127;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      30 ERKEEK 35
Db      107 ERKEEK 112

RESULT 10
US-09-270-767-53607
/ Sequence 53607, Application US/09270767
/ Patent No. 6703491
/ GENERAL INFORMATION:
/ APPLICANT: Homburger et al.
/ TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
/ FILE REFERENCE: File Reference: 7326-094
/ CURRENT APPLICATION NUMBER: US/09/270,767
/ CURRENT FILING DATE: 1999-03-17
/ NUMBER OF SEQ ID NOS: 62517
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 53607
/ LENGTH: 127
/ TYPE: PRT
/ ORGANISM: Drosophila melanogaster
/ FEATURE:
/ OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-53607

Query Match      14.6%; Score 6; DB 4; Length 127;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      30 ERKEEK 35
Db      107 ERKEEK 112

RESULT 11
5229115-2
/ Patent No. 5229115
/ APPLICANT: LYNCH, DAVID H.
/ TITLE OF INVENTION: ADOPTIVE IMMUNOTHERAPY WITH INTERLEUKIN-7
/ NUMBER OF SEQUENCES: 2
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/07/559,001
/ FILING DATE: 26-JUL-1990
/ SEQ ID NO: 2
/ LENGTH: 129
5229115-2

Query Match      14.6%; Score 6; DB 6; Length 129;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 32 KEEKNV 37
| | | | |
DB 94 KEEKNV 99

RESULT 12

US-09-708-200-17
; Sequence 17, Application US/09708200
; Patent No. 6576468
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas C
; APPLICANT: Grasso, Luigi
; APPLICANT: Sasse, Philip M
; TITLE OF INVENTION: METHODS FOR ISOLATING NOVEL ANTIMICROBIAL AGENTS FROM
; FILE REFERENCE: MOR-0005
; CURRENT APPLICATION NUMBER: US/09/708,200
; CURRENT FILING DATE: 2000-11-07
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-708-200-17

Query Match 14.6%; Score 6; DB 4; Length 133;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 LSTAVK 25
| | | | |
DB 35 LSTAVK 40

RESULT 13

US-09-248-796A-22195
; Sequence 22195, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 22195
; LENGTH: 135
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-22195

Query Match 14.6%; Score 6; DB 4; Length 135;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 KNVVGS 40
| | | | |
DB 55 KNVVGS 60

RESULT 14

US-09-710-279-1214
; Sequence 1214, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1214
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-710-279-1214

Query Match 14.6%; Score 6; DB 4; Length 147;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 VKLIAP 29
| | | | |
DB 120 VKLIAP 125

RESULT 15

US-08-446-908-2
; Sequence 2, Application US/08446908
; Patent No. 5705149
; GENERAL INFORMATION:
; APPLICANT: Namen, Anthony E.
; APPLICANT: Goodwin, Raymond G.
; APPLICANT: Lupton, Stephen D.
; APPLICANT: Mochizuki, Diane Y.
; TITLE OF INVENTION: Interleukin-7 and Antibodies Reactive
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: US
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple 7.1
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,908
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/231,205
; FILING DATE: 21-APR-1994
; APPLICATION NUMBER: US 07/957,649
; FILING DATE: 06-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,438
; FILING DATE: 13-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/255,209
; FILING DATE: 07-OCT-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/113,566
; FILING DATE: 26-OCT-1987
; ATTORNEY/AGENT INFORMATION:
; NAME: Seese, Kathryn A.
; REGISTRATION/DOCKET NUMBER: 32,172
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430

TELEFAX: (206) 233-0644
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 154 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-446-908-2

Query Match 14.6% Score 6; DB 1; Length 154;
 Best Local Similarity 100.0%; Pred. No. 99;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 KEEKV 37
 |||||
 Db 119 KEEKV 124

Search completed: February 1, 2005, 15:35:42
 Job time : 18.5556 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: February 1, 2005, 15:09:39 ; Search time 59.7778 Seconds

(without alignments)
247.799 Million cell updates/sec

Title: SE05SER1694

Perfect score: 41
Sequence: 1 LKLLCKNFGAENPDFFVPLV.....TAVKLIAPEKKEKLVLSA 41

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1608061 seqs, 361289386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1311380

Minimum DB seq length: 21

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	48.8	1149	US-10-128-558-167	Sequence 167, App
2	7	17.1	61	US-10-767-701-52614	Sequence 52614, A
3	7	17.1	75	US-10-424-589-22449	Sequence 22449, A
4	7	17.1	79	US-10-424-589-157348	Sequence 157348, A
5	7	17.1	93	US-10-424-599-198515	Sequence 198515, A
6	7	17.1	98	US-10-424-599-271135	Sequence 271135, A
7	7	17.1	119	US-10-424-599-283405	Sequence 283405, A
8	7	17.1	169	US-10-767-701-53909	Sequence 53909, A
9	7	17.1	186	US-10-424-589-279781	Sequence 279781, A
10	7	17.1	232	US-10-424-589-157336	Sequence 157336, A
11	7	17.1	265	US-10-424-599-157334	Sequence 157334, A
12	7	17.1	308	US-10-282-122A-52352	Sequence 52352, A
13	7	17.1	319	US-10-424-599-232288	Sequence 232288, A

14	7	17.1	324	US-10-425-115-323502	Sequence 323502, A
15	7	17.1	332	US-10-425-114-62939	Sequence 62939, A
16	7	17.1	337	US-10-369-493-20250	Sequence 20250, A
17	7	17.1	416	US-10-437-963-109626	Sequence 109626, A
18	7	17.1	534	US-10-389-566-822	Sequence 822, App
19	7	17.1	534	US-10-389-566-1163	Sequence 1163, App
20	7	17.1	574	US-10-437-963-109624	Sequence 109624, A
21	6	14.6	41	US-10-437-963-152060	Sequence 152060, A
22	6	14.6	42	US-09-864-761-47124	Sequence 47124, A
23	6	14.6	44	US-09-864-761-35349	Sequence 35349, A
24	6	14.6	44	US-10-767-701-47684	Sequence 47684, A
25	6	14.6	47	US-09-864-761-40474	Sequence 40474, A
26	6	14.6	47	US-10-371-634-25	Sequence 25, App1
27	6	14.6	50	US-09-764-877-1063	Sequence 1063, App
28	6	14.6	50	US-10-242-515-1063	Sequence 1063, App
29	6	14.6	50	US-10-425-115-253833	Sequence 253833, A
30	6	14.6	53	US-10-425-115-285594	Sequence 285594, A
31	6	14.6	55	US-10-424-599-254164	Sequence 254164, A
32	6	14.6	57	US-10-424-599-166655	Sequence 166655, A
33	6	14.6	57	US-10-425-115-358196	Sequence 358196, A
34	6	14.6	61	US-10-002-344A-231	Sequence 231, App
35	6	14.6	62	US-10-282-122A-48051	Sequence 48051, A
36	6	14.6	64	US-10-425-115-335280	Sequence 335280, A
37	6	14.6	67	US-10-424-599-218416	Sequence 218416, A
38	6	14.6	68	US-10-437-963-169420	Sequence 169420, A
39	6	14.6	69	US-10-425-115-264251	Sequence 264251, A
40	6	14.6	70	US-09-880-192-49	Sequence 49, App1
41	6	14.6	70	US-10-427-348-49	Sequence 49, App1
42	6	14.6	71	US-10-424-599-238496	Sequence 238496, A
43	6	14.6	72	US-10-425-115-264469	Sequence 264469, A
44	6	14.6	74	US-10-424-599-187518	Sequence 187518, A
45	6	14.6	77	US-09-925-300-1880	Sequence 1880, App

ALIGNMENTS

RESULT 1
US-10-128-558-167
Sequence 167, Application US/10128558
Publication No. US20040219521A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Wang, Zhiwei
APPLICANT: Meng, Gezhil
APPLICANT: Boyle, Bryan J
APPLICANT: Drmanac, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
FILE REFERENCE: 812A
CURRENT APPLICATION NUMBER: US/10/128,558
CURRENT FILING DATE: 2002-04-22
PRIOR APPLICATION NUMBER: US 60/339,453
PRIOR FILING DATE: 2001-12-11
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-04-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: PCT/US01/03800
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 09/515,126
PRIOR FILING DATE: 2000-02-28
Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 412
SOFTWARE: pc_fl_genes version 6.0
SEQ ID NO: 167
LENGTH: 1149
TYPE: PRT
ORGANISM: Homo sapiens
US-10-128-558-167

Query Match 48.8%; Score 20; DB 17; Length 1149;
Best Local Similarity 100.0%; Pred. No. 1.8e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LKLLCKNFGAENDDPVPVL 20
DB 732 LKLLCKNFGAENDDPVPVL 751

RESULT 2

US-10-767-701-52614
Sequence 52614, Application US/10767701
Publication No. US20040172684A1
GENERAL INFORMATION:
APPLICANT: Kovacic, David K.
APPLICANT: Zhou, Yihua
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/767,701
CURRENT FILING DATE: 2004-01-29
NUMBER OF SEQ ID NOS: 63128
SEQ ID NO 52614
LENGTH: 61
TYPE: PRT
ORGANISM: Sorghum bicolor
FEATURE:
OTHER INFORMATION: Clone ID: 12691726.pdp
US-10-767-701-52614

Query Match 17.1%; Score 7; DB 16; Length 61;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 APERKEE 34
DB 19 APERKEE 25

RESULT 3

US-10-424-599-224449
Sequence 224449, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J
APPLICANT: Kovacic, David K
APPLICANT: Zhou, Yihua
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 224449
LENGTH: 75
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MFT3847_44706C.1.pdp
US-10-424-599-224449

Query Match 17.1%; Score 7; DB 15; Length 75;
Best Local Similarity 100.0%; Pred. No. 24;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 29 PERKEEK 35
DB 4 PERKEEK 10

RESULT 4

US-10-424-599-157348
Sequence 157348, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J
APPLICANT: Kovacic, David K
APPLICANT: Zhou, Yihua
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 157348
LENGTH: 79
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(79)
OTHER INFORMATION: unsure at all Xaa locations
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MFT3847_113105C.1.pdp
US-10-424-599-157348

Query Match 17.1%; Score 7; DB 15; Length 79;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 PERKEEK 35
DB 25 PERKEEK 31

RESULT 5

US-10-424-599-198515
Sequence 198515, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J
APPLICANT: Kovacic, David K
APPLICANT: Zhou, Yihua
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 198515
LENGTH: 93
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MFT3847_21283C.1.pdp
US-10-424-599-198515

Query Match 17.1%; Score 7; DB 15; Length 93;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 PERKEEK 35
DB 64 PERKEEK 70

```
RESULT 6
US-10-424-599-271135
; Sequence 271135, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 271135
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_86854C.1.pap
US-10-424-599-271135

Query Match          17.1%; Score 7; DB 15; Length 98;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 LKLLCKN 7
        |||||
DB      70 LKLICKN 76

RESULT 7
US-10-424-599-283405
; Sequence 283405, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 283405
; LENGTH: 119
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(119)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_97938C.1.pap
US-10-424-599-283405

Query Match          17.1%; Score 7; DB 15; Length 119;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      32 KEKKNVL 38
        |||||
DB      67 KEKKNVL 73

RESULT 8
US-10-767-701-53909
; Sequence 53909, Application US/10767701
```

```
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 53909
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 13784602.pap
US-10-767-701-53909

Query Match          17.1%; Score 7; DB 16; Length 169;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      28 APERKEE 34
        |||||
DB      29 APERKEE 35

RESULT 9
US-10-424-599-279781
; Sequence 279781, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 279781
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(186)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_94664C.1.pap
US-10-424-599-279781

Query Match          17.1%; Score 7; DB 15; Length 186;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      30 ERKEEKN 36
        |||||
DB      73 ERKEEKN 79

RESULT 10
US-10-424-599-157336
; Sequence 157336, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
```

```
/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
/ FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(53223)B
/ CURRENT APPLICATION NUMBER: US/10/424,599
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 285684
/ SEQ ID NO 157336
/ LENGTH: 232
/ TYPE: PRT
/ ORGANISM: Glycine max
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT3847_113095C.1.pep
US-10-424-599-157336

Query Match          17.1%; Score 7; DB 15; Length 232;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      32 KEKNVL 38
Db      180 KEKNVL 186

RESULT 11
US-10-424-599-157334
/ Sequence 157334, Application US/10424599
/ Publication No. US20040031072A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa Thomas J
/ APPLICANT: Kovalic David K
/ APPLICANT: Zhou Yihua
/ APPLICANT: Cao Yongwei
/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
/ FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(53223)B
/ CURRENT APPLICATION NUMBER: US/10/424,599
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 285684
/ SEQ ID NO 157334
/ LENGTH: 265
/ TYPE: PRT
/ ORGANISM: Glycine max
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT3847_113093C.1.pep
US-10-424-599-157334

Query Match          17.1%; Score 7; DB 15; Length 265;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      32 KEKNVL 38
Db      213 KEKNVL 219

RESULT 12
US-10-282-122A-52352
/ Sequence 52352, Application US/10282122A
/ Publication No. US20040029129A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Karl
/ APPLICANT: Zykkind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Foreyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
```

```
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 52352
/ LENGTH: 308
/ TYPE: PRT
/ ORGANISM: Clostridium botulinum
US-10-282-122A-52352

Query Match          17.1%; Score 7; DB 15; Length 308;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      35 KNVIGSA 41
Db      298 KNVIGSA 304

RESULT 13
US-10-424-599-232288
/ Sequence 232288, Application US/10424599
/ Publication No. US20040031072A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa Thomas J
/ APPLICANT: Kovalic David K
/ APPLICANT: Zhou Yihua
/ APPLICANT: Cao Yongwei
/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
/ FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(53223)B
/ CURRENT APPLICATION NUMBER: US/10/424,599
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 285684
/ SEQ ID NO 232288
/ LENGTH: 319
/ TYPE: PRT
/ ORGANISM: Glycine max
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT3847_5177C.1.pep
US-10-424-599-232288

Query Match          17.1%; Score 7; DB 15; Length 319;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      21 STAVKLI 27
Db      13 STAVKLI 19
```

RESULT 14

US-10-425-115-323502
 ; Sequence 323502, Application US/10425115
 ; Publication No. US20040214272A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE REFERENCE: 38-21(53222)B
 ; CURRENT APPLICATION NUMBER: US/10/425,115
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 369326
 ; SEQ ID NO 323502
 ; LENGTH: 324
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: MRT4577_58103C.1.pep
 US-10-425-115-323502

Query Match

17.1%; Score 7; DB 17; Length 324;

Best Local Similarity 100.0%; Pred. No. 86;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 STAVKLI 27
 |||||
 |||||

DB 3 STAVKLI 9

RESULT 15

US-10-425-114-62939
 ; Sequence 62939, Application US/10425114
 ; Publication No. US2004003488A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Jingdong
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Screen, Steven E
 ; APPLICANT: Tabaska, Jack E
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE REFERENCE: 38-21(53313)B
 ; CURRENT APPLICATION NUMBER: US/10/425,114
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 73128
 ; SEQ ID NO 62939
 ; LENGTH: 332
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: LIB3069-044-D11_FLI.pep
 US-10-425-114-62939

Query Match

17.1%; Score 7; DB 15; Length 332;

Best Local Similarity 100.0%; Pred. No. 88;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 STAVKLI 27
 |||||
 |||||

DB 11 STAVKLI 17

Search completed: February 1, 2005, 15:44:46
 Job time : 60.7778 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:55:01 ; Search time 13.5556 Seconds
(without alignments)
291.016 Million cell updates/sec

Title: SEQ5SER1694
Perfect score: 41
Sequence: 1 LKLLCKNKGAEKNDPFPVFL.....TAVKLIAPERKEKVLGSA 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 279530

Minimum DB seq length: 21
Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	17.1	330	2 T35360	probable membrane
2	7	17.1	435	2 B72418	conserved hypothet
3	7	17.1	506	2 T46196	cytochrome P450-11
4	7	17.1	759	2 F86362	F19G10.17 protein
5	7	17.1	3566	1 A40701	tenascin-X precurs
6	6	14.6	45	2 H81526	hypothetical prote
7	6	14.6	77	2 C70306	conserved hypothet
8	6	14.6	98	2 UC2403	PMS8 homolog miema
9	6	14.6	105	2 B71811	conserved hypothet
10	6	14.6	109	2 B61180	protein-tyrosine-p
11	6	14.6	114	2 T34582	hypothetical prote
12	6	14.6	125	2 S04503	pancreatic ribonuc
13	6	14.6	139	2 AD2997	hypothetical prote
14	6	14.6	141	2 B96286	hypothetical prote
15	6	14.6	144	2 A70411	small heat shock p
16	6	14.6	148	2 T49397	hypothetical prote
17	6	14.6	154	2 S03171	interleukin-7 prec
18	6	14.6	157	2 G72864	Acotf-118 protein
19	6	14.6	159	2 UC2401	PMS6 homolog miema
20	6	14.6	161	2 UC2402	PMS7 homolog miema
21	6	14.6	172	2 DB1719	conserved hypothet
22	6	14.6	186	2 JC2400	PMS5 homolog miema
23	6	14.6	187	2 S26139	signaling protein
24	6	14.6	187	2 A48901	beta-lactamase exp
25	6	14.6	188	2 E83080	hypothetical prote
26	6	14.6	192	2 T19151	hypothetical prote
27	6	14.6	195	2 S75344	hypothetical prote
28	6	14.6	199	2 A11317	hypothetical prote
29	6	14.6	203	2 A11317	hypothetical prote

30	6	14.6	203	2 A11689	hypothetical prote
31	6	14.6	215	2 T23482	hypothetical prote
32	6	14.6	231	2 T12012	cuticular protein
33	6	14.6	249	2 H64368	cobalamin biosynth
34	6	14.6	252	2 JC2399	PMS4 homolog miema
35	6	14.6	256	2 UC2398	PMS3 homolog miema
36	6	14.6	266	2 T19236	hypothetical prote
37	6	14.6	275	2 G90221	conserved hypothet
38	6	14.6	277	2 T40033	probable mitochond
39	6	14.6	279	2 T48013	hypothetical prote
40	6	14.6	282	2 S61860	hpc protein - Pse
41	6	14.6	289	2 AD1086	ATP synthase gamma
42	6	14.6	295	2 S76136	hypothetical prote
43	6	14.6	298	2 T31518	hypothetical prote
44	6	14.6	302	2 A84331	hypothetical prote
45	6	14.6	305	2 A69295	UDP-glucose 4-epim

ALIGNMENTS

RESULT 1
T35360
Probable membrane protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004
C:Accession: T35360
R:Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
Submitted to the EMBL Data Library, June 1999
A:Reference number: Z21576
A:Accession: T35360
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-330 <MDB>
A:Cross-references: UNIPROT:Q9XAL0; EMBL:AL079348; PDB:CB45459.1; GSPDB:GN00070; SCOPD
A:Experimental source: strain A3 (2)
C:Genetics:
A:Gene: SCOPDB:SC6ET3.03

Query Match 17.1%; Score 7; DB 2; Length 330;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 GAENPDP 15
Db 109 GAENPDP 115

RESULT 2
B72418
Conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: B72418
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: B72418
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-435 <ARN>
A:Cross-references: UNIPROT:Q9WXV3; GB:AE001696; GB:AE000512; NID:g4980582; PIDN:AAD519
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0098
C:Superfamily: GTP-binding protein obg; translation elongation factor Tu homology
F165-290/Domains: translation elongation factor Tu homology <ETU>
Query Match 17.1%; Score 7; DB 2; Length 435;
Best Local Similarity 100.0%; Pred. No. 15;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 PERKEEK 35
 |||||
 Db 342 PERKEEK 348

RESULT 3
 cytochrome P450-like protein - Arabidopsis thaliana
 T46196
 N/Alternate names: protein T8P19.30
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
 C/Accession: T46196
 R/Choiene, N.; Robert, C.; Brottier, P.; Mincker, P.; Cartolico, L.; Artigunave, F.; Sa
 submitted to the Protein Sequence Database, December 1999
 A/Reference number: 223008
 A/Accession: T46196
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-506 <CHO>
 A/Cross-references: UNIPROT:O9SWP5; EMBL:AL13315
 A/Experimental source: cultivar Columbia; BAC clone T8P19
 C/Genetics:
 A/Map position: 3
 A/Note: T8P19.30
 C/Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology
 F/300-469/Domain: cytochrome P450 homology <P45>

Query Match 17.1%; Score 7; DB 2; Length 506;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 VPLVSTA 23
 |||||
 Db 160 VPLVSTA 166

RESULT 4
 F86362
 F19G10.17 protein - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
 C/Accession: F86362
 R/Thelouis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Corn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 aisen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A/Authors: Salzeberg, S.L.; Schwartz, J.R.; Shin, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.
 A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A/Reference number: A86141; MUID:21016719; PMID:11130712
 A/Accession: F86362
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-759 <STO>
 A/Cross-references: UNIPROT:O23135; GB:AE005172; NID:g2462837; PIDD:AA87212.1; GSPDB:GN
 C/Genetics:
 A/Map position: 1

Query Match 17.1%; Score 7; DB 2; Length 759;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 VPLVSTA 23
 |||||
 Db 277 VPLVSTA 283

RESULT 5

A40701
 tenascin-X precursor - human
 C/Species: Homo sapiens (man)
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
 C/Accession: A40701; A33725; C42175
 R/Bristow, J.; Tee, M.K.; Gitelman, S.E.; Mellon, S.H.; Miller, W.L.
 J. Cell Biol. 122, 265-278, 1999
 A/Title: Tenascin-X: a novel extracellular matrix protein encoded by the human XB gene o
 A/Reference number: A40701; MUID:93300909; PMID:7686164
 A/Accession: A40701
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-3566 <BRI>
 A/Cross-references: UNIPROT:P22105; UNIPROT:O9UC11; EMBL:X71937
 R/Morel, Y.; Bristow, J.; Gitelman, S.E.; Miller, W.L.
 Proc. Natl. Acad. Sci. U.S.A. 86, 6582-6586, 1989
 A/Title: Transcript encoded on the opposite strand of the human steroid 21-hydroxylase/c
 A/Reference number: A33725; MUID:89367293; PMID:2475872
 A/Accession: A33725
 A/Molecule type: mRNA
 A/Residues: 2748-3199, 'V', 3201-3298, 'E', 3299-3314, 'G', 3316-3566 <MOR>
 A/Cross-references: GB:M25813; NID:g183069; PIDD:AA35884.1; PIDD:g183070
 R/Matsunoto, K.; Arai, M.; Ishihara, N.; Ando, A.; Inoko, H.; Ikemura, T.
 Genomics 12, 485-491, 1992
 A/Title: Cluster of fibronectin type III repeats found in the human major histocompatib
 enascin.
 A/Reference number: A42175; MUID:92217969; PMID:1373119
 A/Accession: C42175
 A/Molecule type: DNA
 A/Residues: 1849-1936 <MAT>
 A/Experimental source: clone 3.9Kf3-1
 A/Note: sequence extracted from NCBI backbone (NCBI:95694)
 C/Genetics:
 A/Gene: GDB:TMXA; D6S103E; TNX; XA; XB
 A/Cross-references: GDB:568487; OMIM:600261
 A/Map position: 6p21.3-6p21.3
 C/Superfamily: tenascin-X; EGF homology; fibronogen beta/gamma homology; fibronectin typ
 C/Keywords: extracellular matrix; glycoprotein
 F/435-461/Domain: EGF homology <EGF>
 F/748-828/Domain: fibronectin type III repeat homology <3F1>
 F/828-856/Domain: fibronectin type III repeat homology <3F2>
 F/873-953/Domain: fibronectin type III repeat homology <3F3>
 F/975-1055/Domain: fibronectin type III repeat homology <3F4>
 F/1078-1158/Domain: fibronectin type III repeat homology <3F5>
 F/1167-1247/Domain: fibronectin type III repeat homology <3F6>
 F/1248-1317/Domain: fibronectin type III repeat homology <3F7>
 F/1323-1403/Domain: fibronectin type III repeat homology <3F8>
 F/1412-1492/Domain: fibronectin type III repeat homology <3F9>
 F/1510-1590/Domain: fibronectin type III repeat homology <3F10>
 F/1618-1676/Domain: fibronectin type III repeat homology <3F11>
 F/1678-1749/Domain: fibronectin type III repeat homology <3F12>
 F/1751-1831/Domain: fibronectin type III repeat homology <3F13>
 F/1849-1929/Domain: fibronectin type III repeat homology <3F14>
 F/1955-2035/Domain: fibronectin type III repeat homology <3F15>
 F/2061-2141/Domain: fibronectin type III repeat homology <3F16>
 F/2167-2246/Domain: fibronectin type III repeat homology <3F17>
 F/2274-2354/Domain: fibronectin type III repeat homology <3F18>
 F/2382-2462/Domain: fibronectin type III repeat homology <3F19>
 F/2488-2568/Domain: fibronectin type III repeat homology <3F20>
 F/2584-2664/Domain: fibronectin type III repeat homology <3F21>
 F/2677-2757/Domain: fibronectin type III repeat homology <3F22>
 F/2771-2851/Domain: fibronectin type III repeat homology <3F23>
 F/2878-2958/Domain: fibronectin type III repeat homology <3F24>
 F/2977-3067/Domain: fibronectin type III repeat homology <3F25>
 F/3078-3159/Domain: fibronectin type III repeat homology <3F26>
 F/3167-3247/Domain: fibronectin type III repeat homology <3F27>
 F/3255-3334/Domain: fibronectin type III repeat homology <3F28>
 F/3349-3357/Domain: fibronogen beta/gamma homology <3F29>

Query Match 17.1%; Score 7; DB 1; Length 3566;
 Best Local Similarity 100.0%; Pred. No. 95;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

A:Cross-references: EMBL:223052; NID:G438141; PIDN:CAA80587.1; PID:G438142
 C:Superfamily: Protein-tyrosine-phosphatase, receptor type alpha; leukocyte common antileukemia antigen; phosphotyrosine phosphatase; receptor; transmembrane protein; tyrosine-kinase; F1-109/Domain: leukocyte common antigen cytosolic domain homology (fragment) <LAC>

Query Match 14.6%; Score 6; DB 2; Length 109;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 ERKEK 35
 |||||
 DB 14 ERKEK 19

RESULT 11

T34582
 hypothetical protein SC10A5.15 SC10A5.15 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor
 C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004
 C:Accession: T34582

R:Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, January 1998
 A:Reference number: 221548

A:Accession: T34582
 A:Status: preliminary; translated from GB/EMBL/DDJ
 A:Molecule type: DNA

A:Residues: 1-114 <MUR>
 A:Cross-references: UNIPROT:O54106; EMBL:AL021529; PIDN:CAA16447.1; GSPDB:GN00070; SCOPED:SC0EDB:SC10A5.15

A:Gene: SC0EDB:SC10A5.15

Query Match 14.6%; Score 6; DB 2; Length 114;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 AVKIA 28
 |||||
 DB 107 AVKIA 112

RESULT 12

S04503
 pancreatic ribonuclease (EC 3.1.27.5) - Ehrenberg's mole-rat

C:Species: Spalax leucodon ehrenbergi (Ehrenberg's mole-rat)
 C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 09-Jul-2004
 C:Accession: S04503; A32505

R:Schueler, C.; Neuboom, B.; Wuebbels, G.H.; Beintema, J.J.; Nevo, E.
 Biol. Chem. Hoppe-Seyler 370, 583-589, 1989
 A:Title: The amino-acid sequence of pancreatic ribonuclease from the mole rat Spalax ehrenbergi
 A:Reference number: S04503; MUID:89374807; PMID:2673297

A:Accession: S04503
 A:Molecule type: protein
 A:Residues: 1-125 <SCH>

A:Cross-references: UNIPROT:P16414
 A>Note: the source is designated as Spalax ehrenbergi
 C:Superfamily: pancreatic ribonuclease
 C:Keywords: glycoprotein; hydrolase

F:12,42,120/Active site: His, Lys, His #status predicted
 F:27-85,41-96,59-111,66-73/Denulfide bonds: #status predicted

Query Match 14.6%; Score 6; DB 2; Length 125;
 Best Local Similarity 100.0%; Pred. No. 56;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 DFVPL 19
 |||||
 DB 114 DFVPL 119

RESULT 13

AD2997
 hypothetical protein Atu3582 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004

C:Accession: AD2997

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
 erag, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavina, T.; Levy, R.; Li, M.; McCelliff
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.
 ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; MUID:21608550; PMID:11743193

A:Accession: AD2997

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-139 <KUR>
 A:Cross-references: UNIPROT:Q8U923; GB:AE008689; PIDN:AAU44394.1; PID:G17741992; GSPDB:GN

A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: Atu3582
 A:Map position: linear chromosome

Query Match 14.6%; Score 6; DB 2; Length 139;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 PFVPL 20
 |||||
 DB 96 PFVPL 101

RESULT 14

E98286
 hypothetical protein AGR_L2491 [imported] - Agrobacterium tumefaciens (strain C58, Cere

C:Species: Agrobacterium tumefaciens
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
 C:Accession: E98286

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; MUID:21608551; PMID:11743194

A:Accession: E98286

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-141 <KUR>

A:Cross-references: UNIPROT:Q8U923; GB:AE007870; PIDN:AAK69815.1; PID:G15159745; GSPDB:G

C:Genetics:
 A:Gene: AGR_L2491

A:Map position: linear chromosome

Query Match 14.6%; Score 6; DB 2; Length 141;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 PFVPL 20
 |||||
 DB 98 PFVPL 103

RESULT 15

A70411
 small heat shock protein (class I) - Aquifex aeolicus

C:Species: Aquifex aeolicus
 C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
 C:Accession: A70411

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ovi

V. Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: A70411

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A;Residues: 1-144 <ADP>
A;Cross-references: UNIPROT:O67316; GB:AE000732; NID:G2983704; PIDN:AAC07270.1; PID:G298
A;Experimental source: strain VFS
C;Genetics:
A;Gene: hspC
C;Superfamily: alpha-crystallin-related small heat shock protein

Query Match 14.6%; Score 6; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 ERKEEK 35
Db 80 ERKEEK 85

Search completed: February 1, 2005, 15:33:00
Job time : 19.5556 secs

This Page Blank (uspto)


```

QY 1 LKLLCKNFGAENPDFFVPLSTAVKLIAPERKEKNVLGSA 41
DB 427 LKLLCKNFGAENPDFFVPLSTAVKLIAPERKEKNVLGSA 467

RESULT 2
BP28 MACFA STANDARD; PRT; 958 AA.
AC 09GM44;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Protein BAP28 (OmpA-17571) (Fragment).
GN Name=BAP28;
OS Maccaca fascicularis (Crap eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Maccaca.
OC NCBI_TaxID=9541;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RT Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.";
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: Belongs to the BAP28 family.
CC -1- SIMILARITY: Contains 1 HEAT repeat.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC
CC EMBL; AB049842; BAB16728.1; ALT_INIT.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000357; HEAT.
DR PROSITE; PS50077; HEAT_REPEAT; FALSE_NEG.
FT NON TER 1 956
FT REPEAT 920 956 HEAT.
SQ SEQUENCE 958 AA; 108644 MW; 3DBD95C3623CFB31 CRC64;

Query Match 100.0%; Score 41; DB 1; Length 958;
Best Local Similarity 100.0%; Pred. NO. 5.4e-34;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LKLLCKNFGAENPDFFVPLSTAVKLIAPERKEKNVLGSA 41
DB 487 LKLLCKNFGAENPDFFVPLSTAVKLIAPERKEKNVLGSA 527

RESULT 3
BP197 PRELIMINARY; PRT; 1106 AA.
AC 06P197;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE FLJ10359 protein (Fragment).
GN Name=FLJ10359;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OC NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RT Krzywinski M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E.,
RL Jones S.J., Maira M.A.;
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RT Krzywinski M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E.,
RL Jones S.J., Maira M.A.;
[2]

```

```

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshitsuki S., Carninci P., Prange C.,
RA Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bonak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Maira M.A.;
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA Strausberg R.L.,
RL Submitted (JAN-2004) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC065205; AAH65205.1;
DR InterPro; IPR008938; ARM.
FT NON TER 1 125359
SQ SEQUENCE 1106 AA; 125359 MW; 09F7CE94042302C4 CRC64;

Query Match 100.0%; Score 41; DB 2; Length 1106;
Best Local Similarity 100.0%; Pred. NO. 6.1e-34;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LKLLCKNFGAENPDFFVPLSTAVKLIAPERKEKNVLGSA 41
DB 636 LKLLCKNFGAENPDFFVPLSTAVKLIAPERKEKNVLGSA 676

RESULT 4
AAH65205 PRELIMINARY; PRT; 1106 AA.
ID AAH65205;
AC AAH65205;
DT 02-MAR-2004 (TREMBlrel. 27, Created)
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)
DE FLJ10359 protein (Fragment).
GN Name=FLJ10359;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OC NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshitsuki S., Carninci P., Prange C.,
RA Raha S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Maira M.A.;
[2]

```

"Generation and initial analysis of more than 15,000 full-length human RT and mouse cDNA sequences.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2] SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RA Straubeberg R.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC065205; AAH65205.1;
 FT NON TIR
 SQ SEQUENCE 1106 AA; 125359 MW; 09F7CE94042302C4 CRC64;
 Query Match 100.0%; Score 41; DB 2; Length 1106;
 Best Local Similarity 100.0%; Pred. No. 6,1e-34;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LKLCNKGAEPPDPFVPLSTAVKLIAPERKEKNVLGSA 41
 DB 636 LKLCNKGAEPPDPFVPLSTAVKLIAPERKEKNVLGSA 676
 RESULT 5
 BP28 HUMAN STANDARD; PRT; 2144 AA.
 ID BP28 HUMAN
 AC O9H583; O9NM23;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Protein BAP28.
 GN Name=BAP28;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANTS SER-1694; ALA-1854; ASP-1967 AND
 RP GXY-2017.
 RA Bouguieret L., Chumakov I., Barry C., Cohen-Akente A.;
 RT "A novel BAP28 gene and protein."
 RL Patent number WO0100659, 04-JAN-2001.
 RN [2]
 RP SEQUENCE OF 1534-2144 FROM N.A.
 RA Cobley V.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1777-2144 FROM N.A.
 RX PubMed:14702039; DOI=10.1038/ng1285;
 RA Oca T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
 RA Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
 RA Shiratori A., Sudo H., Hosokita T., Kaku Y., Kodaira H., Kondo H.,
 RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
 RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
 RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
 RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y.,
 RA Iehida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Houchi T.,
 RA Kuwano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
 RA Nomura Y., Togai S., Komai F., Hara R., Takeuchi K., Arita M.,
 RA Imose N., Mutsaers K., Yuuki H., Oshima A., Sasaki N., Aotsuma S.,
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
 RA Moriya S., Komiyama H., Sato H., Takami S., Terashima Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 RA Hishigaki H., Watanabe K., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirose M., Ohmori Y.,
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsunuma K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
 RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,

RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuko Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
 RT "Complete sequencing and characterization of 21,243 full-length human
 RT cDNAs";
 RL Nat. Genet. 36:40-45(2004).
 CC -1- SIMILARITY: Belongs to the BAP28 family.
 CC -1- SIMILARITY: Contains 1 HEAT repeat.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AX067150; CAC26776.1; -
 DR EMBL: AL136105; CAC15948.1; -
 DR EMBL: AK001221; BAA91564.1; ALT_INIT.
 DR SWISS-2DPAGE: O9H583; HUMAN.
 DR InterPro: IPR008938; ARM.
 DR InterPro: IPR000357; HEAT.
 DR PROSITE: PS50077; HEAT_REPEAT; FALSE_NEG.
 KW Polymorphism.
 FT REPEAT 2106 2142 HEAT.
 FT VARIANT 1694 1694 N -> S.
 FT VARIANT 1854 1854 V -> A.
 FT VARIANT 1967 1967 N -> D.
 FT VARIANT 2017 2017 E -> G.
 FT VARIANT 2017 2017 /FTID=VAR_010941.
 FT SEQUENCE 2144 AA; 242355 MW; D6815E78D8C8B7 CRC64;
 Query Match 48.8%; Score 20; DB 1; Length 2144;
 Best Local Similarity 100.0%; Pred. No. 9.8e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LKLCNKGAEPPDPFVPL 20
 DB 1674 LKLCNKGAEPPDPFVPL 1693
 RESULT 6
 ID Q9J877 PRELIMINARY; PRT; 114 AA.
 AC Q9J877;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE ORF58.
 OS Spodoptera exigua nucleopolyhedrovirus.
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 OC Nucleopolyhedrovirus.
 OX NCBI_TaxID=10454;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20036646; PubMed=10567663;
 RA IJkel W.F., van Strien E.A., Heldens J.G., Broer R., Zuidema D.,
 RA Goldbach R.W., Vlak J.M.;
 RT "Sequence and organization of the spodoptera exigua multicausid
 RT nucleopolyhedrovirus genome."
 RL J. Gen. Virol. 80:3289-3304(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA IJkel W.F.J., van Strien E.A., Heldens J.G.M., Broer R., Zuidema D.,
 RA Goldbach R.W., Vlak J.M.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF169823; AAF33588.1; -
 DR InterPro: IPR009133; Baculo_11_kDa.
 DR Pfam: PF06143; Baculo_11_kDa; 1.
 SQ SEQUENCE 114 AA; 12644 MW; 8E5AC16B14982C CRC64;

Query Match 17.1%; Score 7; DB 2; Length 114;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 LSTAVKL 26
 Db 65 LSTAVKL 71

RESULT 7

06DJT5 PRELIMINARY; PRT; 209 AA.
 ID 06DJT5
 AC 06DJT5;
 DT 01-OCT-2004 (TREMBlrel. 28, Created)
 DT 01-OCT-2004 (TREMBlrel. 28, Last sequence update)
 DT 01-OCT-2004 (TREMBlrel. 28, Last annotation update)
 DE Hypothetical protein.
 OS Xenopus laevis (African clawed frog).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 CC Xenopodinae; Xenopus.
 CX NCBI_TaxID=8355;

RA [1]
 RA SEQUENCE FROM N.A.
 RA TISSUE=Kidney;
 RA MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Dichtenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stappleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshitsuki S., Carrinci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bock S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.W., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Heltan E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Kravitski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E.,
 RA Jones S.U., Marra M.A.,
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]

RP SEQUENCE FROM N.A.
 RP TISSUE=Kidney;
 RP MEDLINE=22341132; PubMed=12454917;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.,
 RA "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 RT initiative."
 RL Dev. Dyn. 225:384-391 (2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RP TISSUE=Kidney;
 RP Klein S., Strausberg R.,
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL, BC075194, AA075194.1, -.
 KW Hypothetical protein.
 SQ SEQUENCE 209 AA; 23557 MW; B52BF04AB345B389 CRC64;

Query Match 17.1%; Score 7; DB 2; Length 209;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 KNFGAEN 12
 Db 156 KNFGAEN 162

RESULT 8
 ISPE CLOTE
 ID ISPE CLOTE STANDARD; PRT; 280 AA.
 AC 0899A2;
 DT 01-OCT-2004 (Rel. 45, Created)
 DT 01-OCT-2004 (Rel. 45, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.146) (CMK)
 DE (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).
 GN Name=Ispe; OrderedLocNames=CTC00283;
 OS Clostridium tetani.
 CC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 CC Clostridium.
 CX NCBI_TaxID=1513;

RA [1]
 RA SEQUENCE FROM N.A.
 RA STRAIN=Massachusetts / E88;
 RA MEDLINE=22457253; PubMed=12552129; DOI=10.1073/pnas.0335853100;
 RA Bruggemann H., Baumeister S., Fricke W.F., Wieser A., Liesegang H.,
 RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
 RA Gottschalk G.;
 RA "The genome sequence of Clostridium tetani, the causative agent of
 RT tetanus disease."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321 (2003).
 CC -1- FUNCTION: Catalyzes the phosphorylation of the position 2 hydroxy
 CC group of 4-diphosphocytidylyl-2C-methyl-D-erythritol (By
 CC similarity).
 CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
 CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-
 CC D-erythritol.
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth
 CC step.
 CC -1- SIMILARITY: Belongs to the GMP kinase family. Ispe subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).

CC EMBL; AEO15936; AAC34927.1; -.
 DR HAMAP; MF 00061; -1.
 DR InterPro; IPR006204; GMP_kinase.
 DR InterPro; IPR000870; Homoser_kin.
 DR InterPro; IPR004424; Ispe.
 DR Pfam; PF00288; GMP_kinases; 1.
 DR PRINTS; PR00958; HOMSERKINASE.
 DR TIGRFAMs; TIGR00154; Ispe; 1.
 KW ATP-binding; Complete proteome; Isoprene biosynthesis; Kinase;
 KM Transferrase.
 FT NP BIND 91 101 ATP (potential).
 FT ACT_SITE 8 8 By similarity.
 FT ACT_SITE 133 133 By similarity.
 SQ SEQUENCE 280 AA; 31616 MW; B132AC792B8CDD7F CRC64;

Query Match 17.1%; Score 7; DB 1; Length 280;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LKLLCKN 7
 Db 204 LKLLCKN 210

RESULT 9
 ID 072JR4 PRELIMINARY; PRT; 284 AA.
 AC 072JR4;
 DT 05-JUL-2004 (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)

QY 6 KNFGAEN 12
 Db 156 KNFGAEN 162

```

DT 05-JUN-2004 (TReMBLrel. 27, last annotation update)
DE Cell division protein ftsX.
GN Name=ftsX; OrderedLocustNames=TTCC0704;
OS Thermus thermophilus (strain HB27 / ATCC BAA-163 / DSM 7039).
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=262724;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15064768;
RA Henne A., Brueggemann H., Raasch C., Wierer A., Hartach T.,
RA Liesegang H., Johann A., Lienard T., Gohl O., Martinez-Arias R.,
RA Jacobl C., Starkvienne V., Schlenceck S., Dencker S., Huber R.,
RA Klenk H.-P., Kramer W., Merkl R., Gottschalk G., Fritz H.-J.;
RT "The genome sequence of the extreme thermophile Thermus
RT thermophilus.";
RL Nat. Biotechnol. 22:547-553 (2004).
DR EMBL; AE017303; AAS81052.1; -.
DR InterPro; IPR003838; DUF214.
DR Pfam; PF02687; FtsX; 1.
KW Cell division; Complete proteome.
SQ SEQUENCE 284 AA; 31042 MW; 36FAC9F453CA7FAB CRC64;

Query Match 17.1%; Score 7; DB 2; Length 284;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 PFVPLVS 21
DB 244 PFVPLVS 250

RESULT 10
AAS81052 PRELIMINARY; PRT; 284 AA.
ID AAS81052;
AC AAS81052;
DT 14-APR-2004 (TReMBLrel. 27, Created)
DT 14-APR-2004 (TReMBLrel. 27, Last sequence update)
DT 11-MAY-2004 (TReMBLrel. 27, Last annotation update)
DE Cell division protein ftsX.
GN FtsX OR TTCC0704.
OS Thermus thermophilus (strain HB27 / ATCC BAA-163 / DSM 7039).
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=262724;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15064768;
RA Henne A., Brueggemann H., Raasch C., Wierer A., Hartach T.,
RA Liesegang H., Johann A., Lienard T., Gohl O., Martinez-Arias R.,
RA Jacobl C., Starkvienne V., Schlenceck S., Dencker S., Huber R.,
RA Klenk H.-P., Kramer W., Merkl R., Gottschalk G., Fritz H.-J.;
RT "The genome sequence of the extreme thermophile Thermus
RT thermophilus.";
RL Nat. Biotechnol. 22:547-553 (2004).
DR EMBL; AE017303; AAS81052.1; -.
KW Cell division.
SQ SEQUENCE 284 AA; 31042 MW; 36FAC9F453CA7FAB CRC64;

Query Match 17.1%; Score 7; DB 2; Length 284;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 PFVPLVS 21
DB 244 PFVPLVS 250

RESULT 11
Q9XALO PRELIMINARY; PRT; 330 AA.
ID Q9XALO;
AC Q9XALO;
DT 01-NOV-1999 (TReMBLrel. 12, Created)

```

```

DT 01-NOV-1999 (TReMBLrel. 12, last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, last annotation update)
DE Putative membrane protein.
GN OrderedLocustNames=SCO3592; ORFNames=SC66T3_03;
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=A3(2) / M145;
RA MEDLINE=21996410; PubMed=12000953; DOI=10.1038/417141a;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.B., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L.D., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K.M., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147 (2002).
DR EMBL; AL939117; CAB45459.1; -.
DR PIR; T35360; T35360.
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
KW Complete proteome.
SQ SEQUENCE 330 AA; 34963 MW; 94CDE924DF1B740F CRC64;

Query Match 17.1%; Score 7; DB 2; Length 330;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GAENDDP 15
DB 109 GAENDDP 115

RESULT 12
Q8TSU2 PRELIMINARY; PRT; 334 AA.
ID Q8TSU2;
AC Q8TSU2;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein MA0703.
GN OrderedLocustNames=MA0703;
OS Methanosaarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C2A / ATCC 35395 / DSM 2834;
RA MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
RA Galagan J.E., Nussbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atcnor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Deatellano K., Johnson R.,
RA Litton L., McSwan P., McKernan K., Talamas J., Tirelli A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame J.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Maccario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542 (2002).
DR EMBL; AE010731; AAM04143.1; -.
GO; GO:0003677; F:DNA binding; IEA.

```

DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR InterPro; IPR003141; Pesterase_PHP_N.
 DR Pfam; PF02231; PHP_N; 1.
 DR SMART; SM00481; POLI1AC; 1.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 334 AA; 38654 MW; 1BCACB4752B316 CRC64;

Query Match 17.1%; Score 7; DB 2; Length 334;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KILAPER 31
 DB 11 KILAPER 17

RESULT 13
 ID Q6ING9 PRELIMINARY; PRT; 373 AA.
 AC Q6ING9;
 DT 05-JUL-2004 (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE MGC82621 protein.
 GN Name=MGC82621;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strusberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=22341132; PubMed=12454917;
 RA Klein S.L., Strusberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.;
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 RT initiative";
 RL Dev. Dyn. 225:384-391 (2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RA Klein S., Strusberg R.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Belongs to the Pp2C family.
 DR EMBL; BC072312; AAH72312.1; -
 DR InterPro; IPR000222; PP2C.
 DR InterPro; IPR001932; PP2C-like.

DR Pfam; PF00481; PP2C; 1.
 DR SMART; SM00332; PP2C; 1.
 DR SMART; SM00331; PP2C SIG; 1.
 DR PROSITE; PS01032; PP2C; 1.
 KW Hydroxylase; Magnesium.
 SQ SEQUENCE 373 AA; 41536 MW; 58713D1A52099BB7 CRC64;

Query Match 17.1%; Score 7; DB 2; Length 373;
 Best Local Similarity 100.0%; Pred. No. 89;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 PERKEK 35
 DB 229 PERKEK 235

RESULT 14
 ID AAH72312 PRELIMINARY; PRT; 373 AA.
 AC AAH72312;
 DT 01-JUN-2004 (TREMBlrel. 27, Created)
 DT 01-JUN-2004 (TREMBlrel. 27, Last sequence update)
 DT 01-JUN-2004 (TREMBlrel. 27, Last annotation update)
 DE Hypothetical protein.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=22341132; PubMed=12454917;
 RA Klein S.L., Strusberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.;
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 RT initiative";
 RL Dev. Dyn. 225:384-391 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strusberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RA Klein S., Strusberg R.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC072312; AAH72312.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 373 AA; 41536 MW; 58713D1A52099BB7 CRC64;

Query Match 17.1%; Score 7; DB 2; Length 373;
 Best Local Similarity 100.0%; Pred. No. 89;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 29 PERKEEK 35
 |||||
 Db 229 PERKEEK 235

RESULT 15

Q9WXV3 PRELIMINARY; PRT; 435 AA.
 ID Q9WXV3
 AC Q9WXV3
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein.
 GN OrderedLocustNames=TM0098;
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 NX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109 / ATCC 43589;
 RX MEDLINE=99287316; PubMed=10360571; DOI=10.1038/20601;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.U.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L.A., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.L.,
 RA Heidelberg J.F., Sutton G.C., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of Thermotoga maritima."
 RL Nature 399:323-329(1999).
 DR EMBL; AE001696; AAD35192.1; -.
 DR PIR; B72418; B72418.
 DR HSRP; P20964; ILNZ.
 DR TIGR; TM0098; -.
 DR GO; GO:0005525; F:GTP binding; IEA.
 DR InterPro; IPR006074; GTP1_OBG_dom.
 DR InterPro; IPR006073; GTP1_OBG_
 DR InterPro; IPR006169; GTP1_OBG_sub.
 DR InterPro; IPR005225; Small_GTP.
 DR Pfam; PF01018; GTP1_OBG; 1.
 DR PRINTS; PR00326; GTP1_OBG.
 DR TIGRFAMs; TIGR00231; small_GTP; 1.
 DR PROSITE; PS00905; GTP1_OBG; 1.
 KM Complete proteome; Hypothetical protein.
 SQ SEQUENCE 435 AA; 48595 MW; 45A496703A4F74D2 CRC64;

Query Match 17.1%; Score 7; DB 2; Length 435;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 29 PERKEEK 35
 |||||
 Db 342 PERKEEK 348

Search completed: February 1, 2005, 15:30:48
 Job time : 73.5556 secs

This Page Blank (uspto)

CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention

XX SQ Sequence 349 AA;

Query Match 100.0%; Score 41; DB 4; Length 349;
 Best Local Similarity 100.0%; Pred. No. 5.5e-36;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKWKKNHMGPFMSILOEHIGVKKKELTSHQSOLTAFFLEA 41
 DB 39 EKWKKNHMGPFMSILOEHIGVKKKELTSHQSOLTAFFLEA 79

RESULT 2
 ID ADE08012 standard; protein; 1149 AA.
 XX
 AC ADE08012;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Novel protein (useful for identifying genetic disorders) #167.
 XX
 KM novel gene; novel protein; tissue marker; molecular weight marker;
 XX chromosome marker; genetic disorder.
 OS Unidentified.
 XX
 PN WO2003054152-A2.
 XX
 PD 03-JUL-2003.
 XX
 PF 10-DEC-2002; 2002WO-US039555.
 XX
 PR 10-DEC-2001; 2001US-0339739P.
 XX
 PR 11-DEC-2001; 2001US-0339453P.
 XX
 PR 14-MAR-2002; 2002US-0365091P.
 XX
 PR 14-MAR-2002; 2002US-0365384P.
 XX
 PR 12-APR-2002; 2002US-0372381P.
 XX
 PR 12-APR-2002; 2002US-0372615P.
 XX
 PR 22-APR-2002; 2002US-00128558.
 XX
 PR 24-APR-2002; 2002US-0376045P.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YN, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
 PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
 PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ,
 XX
 DR MPI; 2003-569235/53.
 XX
 DR N-PSDB; ADE07101.
 XX
 PT New polynucleotides, useful for expressing recombinant proteins for
 PT analysis, characterization or therapeutic use, or as markers for tissues
 PT in which the corresponding protein is preferentially expressed.
 XX
 PS Claim 20; SEQ ID NO 1078; 1177bp; English.
 XX
 CC The invention comprises the amino acid and coding sequences of novel
 CC proteins. The DNA and protein sequences of the invention are useful as:

CC markers for tissues in which the corresponding protein is preferentially
 CC expressed; as molecular weight markers on gels; as chromosome markers or
 CC tags; to identify chromosomes or to map related gene positions; and to
 CC compare with endogenous DNA sequences in patients to identify potential
 CC genetic disorders. The present amino acid sequence represents a protein
 CC of the invention.

XX SQ Sequence 1149 AA;

Query Match 87.8%; Score 36; DB 7; Length 1149;
 Best Local Similarity 100.0%; Pred. No. 3.8e-30;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 NHMGPFMSILOEHIGVKKKELTSHQSOLTAFFLEA 41
 DB 844 NHMGPFMSILOEHIGVKKKELTSHQSOLTAFFLEA 879

RESULT 3
 ID AAW54099 standard; protein; 515 AA.
 XX
 AC AAW54099;
 XX
 DT 28-SEP-1998 (first entry)
 XX
 DE Homo sapiens BAP28 sequence.
 XX
 KM BARD1; ring protein; BRCA1; breast cancer; risk; diagnosis.
 XX
 OS Homo sapiens.
 XX
 PN WO9812327-A2.
 XX
 PD 26-MAR-1998.
 XX
 PF 19-SEP-1997; 97WO-US016842.
 XX
 PR 20-SEP-1996; 96US-0025296P.
 XX
 PR 03-APR-1997; 97US-0042611P.
 XX
 PR 04-APR-1997; 97US-0042985P.
 XX
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Bowcock AM, Baer R;
 PI
 DR MPI; 1998-230317/20.
 XX
 DR N-PSDB; AAV24135.
 XX
 PT DNA sequence encoding BARD1, B123, BE2, BE14, BE31 or BE445 - which as
 PT breast cancer antigen, BRCA1, binding proteins are useful to identify
 PT patient having or at risk of developing cancer.
 XX
 PS Disclosure; Page 287-288; 348pp; English.
 XX
 CC The sequence is that of a protein which can be used in the preparation of
 CC the recombinant breast cancer antigen, BRCA1, binding proteins BARD1,
 CC B123, BE2, BE14, BE31 or BE445, or a composition for the detection of a
 CC BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid sequence, specifically
 CC a wild type BARD1 composition for the detection or purification of BRCA1,
 CC useful to identify a patient having, or at risk of developing cancer.
 CC BARD1 can be used in the preparation of an anti-BARD1 antibody, and in
 CC the detection and purification of a BRCA1 protein. BARD1, B123, BE2,
 CC BE14, BE31 or BE445 can be used in the identification of a binding protein
 CC agonist or antagonist that alters the binding of BARD1, B123, BE2, BE14,
 CC BE31 or BE445 to BRCA1 or the biological activity of the BRCA1-BARD1,
 CC B123, BE2, BE14, BE31 or BE445 complex. The antibodies can be used to
 CC detect BARD1, B123, BE2, BE14, BE31 or BE445, a specific anti-BARD1
 CC antibody can be used to identify a patient having or at risk of
 CC developing cancer

XX SQ Sequence 515 AA;

Query Match 48.8%; Score 20; DB 2; Length 515;
 Best Local Similarity 100.0%; Pred. No. 3.3e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 EKXWKNHGPFMSILOEHIG 20
 205 EKXWKNHGPFMSILOEHIG 224

RESULT 4

AAB85029
 ID AAB85029 standard; protein, 2144 AA.

AC AAB85029;

DT 06-AUG-2001 (first entry)

DE Protein encoded by BAP28 cDNA consisting of exons 1 to 45.

XX BAP28; prostate; tumour; cancer; diagnostic; genetic analysis.

XX Homo sapiens.

XX Key Location/Qualifiers

XX MISC-difference 1694

XX MISC-difference /label= Ser or Asn

XX MISC-difference 1854

XX MISC-difference /label= Ala or Val

XX MISC-difference 1967

XX MISC-difference /label= Asp or Asn

XX MISC-difference 2017

XX MISC-difference /label= Gly or Glu

XX WO200100669-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-IB001183.

XX 25-JUN-1999; 98US-0141323P.

XX 18-JAN-2000; 2000US-0176880P.

XX (GEST) GENSET.

XX Barry C, Bougueteleret L, Chumakov I, Cohen-Akenine A;

XX WPI; 2001-367032/38.

XX N-PSDB; AAF83909, AAF83910.

XX New BAP28 polynucleotides and polypeptides overexpressed in prostate

XX cancer cells for diagnosing prostate tumors, e.g. by hybridization or

XX polymerase chain reaction assays.

XX Claim 14; Page 297-304; 349pp; English.

XX The invention is directed to BAP28 polypeptides, BAP28 polynucleotide

XX sequences and regulatory region located at the 3' and 5' ends of the

XX BAP28 coding region. The BAP28 polypeptides can be expressed by standard

XX recombinant methodology. BAP28 polynucleotides and polypeptides have been

XX found to be over expressed in prostate tumour cells, therefore levels of

XX BAP28 expression and/or activity may be assayed (e.g. by polymerase chain

XX reaction (PCR)) to diagnose patient suffering from or susceptible to

XX prostate cancer. Antibodies specific for the BAP28 polypeptides are

XX useful as diagnostic reagents. Biallelic markers of the BAP28 gene are

XX useful in genetic analysis. The present sequence represents a protein

XX encoded by a first cDNA sequence of the BAP28 gene consisting of the

XX exons 1 to 45

XX Sequence 2144 AA;

XX Query Match 48.8%; Score 20; DB 4; Length 2144;

XX Best Local Similarity 100.0%; Pred. No. 1.3e-12;

XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 EKXWKNHGPFMSILOEHIG 20
 1834 EKXWKNHGPFMSILOEHIG 1853

RESULT 5

ABO61395
 ID ABO61395 standard; protein, 411 AA.

AC ABO61395;

DT 29-JUL-2004 (first entry)

DE Klebsiella pneumoniae polypeptide seqid 7912.

XX Recombinant expression vector; transcription regulatory element;

XX Klebsiella pneumoniae protein; antibacterial; Vaccine.

XX Klebsiella pneumoniae.

XX US6610836-B1.

XX 26-AUG-2003.

XX 27-JAN-2000; 2000US-00489039.

XX 29-JAN-1999; 99US-0117747P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Breton GL, Osborne M;

XX WPI; 2003-895346/82.

XX N-PSDB; ACH94946.

XX New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for

XX preparing a vaccine composition against Klebsiella pneumoniae.

XX Disclosure; SEQ ID NO 7912; 932pp; English.

XX The invention describes a new isolated nucleic acid encoding a Klebsiella

XX pneumoniae polypeptide. Also described are: a recombinant expression

XX vector comprising the nucleic acid, operably linked to a transcription

XX regulatory element; and a cell comprising the recombinant expression

XX vector. The nucleic acid is useful for preparing a vaccine composition

XX against Klebsiella pneumoniae. This is the amino acid sequence of a

XX Klebsiella pneumoniae polypeptide of the invention

XX Sequence 411 AA;

XX Query Match 19.5%; Score 8; DB 7; Length 411;

XX Best Local Similarity 100.0%; Pred. No. 3.1;

XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 34 LTAFLEA 41

XX 46 LTAFLEA 53

XX RESULT 6

XX ABO42163

XX ID ABO42163 standard; protein, 443 AA.

XX AC ABO42163;

XX 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #27690.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Pseudomonas syringae.

XX XX
 PN WO200271183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342823P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELITR-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;
 PI Wall D, Trawick JD, Cair GJ, Yamamoto R, Forsyth RA, Xu HH;
 DR WPI; 2003-029926/02.
 DR N-PSDB; ACA46033.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 70087; 1766bp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 443 AA;
 Query Match 19.5%; Score 8; DB 6; Length 443;
 Best Local Similarity 100.0%; Pred. No. 3.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 34 LTAFFLEA 41
 |||||
 DB 67 LTAFFLEA 74

XX XX
 AC ABM69250;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Photorhabdus luminescens protein sequence #2347.
 XX
 KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 KW whooping cough.
 XX
 OS Photorhabdus luminescens.
 XX
 PN WO200294867-A2.
 XX
 PD 28-NOV-2002.
 XX
 PF 07-FEB-2002; 2002WO-IB003040.
 XX
 PR 07-FEB-2001; 2001FR-00001659.
 XX
 PA (INSP) INST PASTER.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kuner F, Danchin A;
 PI Buchrieser C;
 PI
 DR WPI; 2003-148459/14.
 XX
 PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
 XX
 PS Claim 2; SEQ ID NO 2347; 1205bp; French.
 XX
 CC The invention relates to the isolation of genes and their encoded
 CC proteins from Photorhabdus luminescens. The isolated sequences are
 CC sources of probes and primers for detecting the genome of *P. luminescens*
 CC and related species; to study polymorphisms; for gene analysis and for
 CC detection/amplification of the genes. Antibodies (Ab) raised against the
 CC polypeptides encoded by the genes are used for detection/identification
 CC of *P. luminescens*, e.g. in foods. The genes, proteins, Ab and cells that
 CC carry a gene-containing vector are used to select compounds that
 CC modulate, regulate, induce or inhibit expression of the genes in plants,
 CC animals or microorganisms other than *P. luminescens* and are able to alter
 CC response or sensitivity to toxins and antibiotics produced by *P.*
 CC luminescens. Cells transformed to express the genes are useful for
 CC recombinant production of the proteins, particularly toxins and
 CC antibacterials useful as insecticides, bactericides and fungicides. The
 CC genes, proteins, vectors containing the genes and Ab are also useful
 CC therapeutically (to treat microbial infection by bacteria or fungi that
 CC are sensitive to *P. luminescens*-encoded toxins or antibiotics) and as
 CC biopesticides. Other uses of the genes and the proteins are as virulence
 CC factors and for identifying targets of human diseases for which *P.*
 CC luminescens is a model (particularly plague and whooping cough). This
 CC sequence represents one of the isolated *P. luminescens* proteins
 XX
 SQ Sequence 446 AA;
 Query Match 19.5%; Score 8; DB 6; Length 446;
 Best Local Similarity 100.0%; Pred. No. 3.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 34 LTAFFLEA 41
 |||||
 DB 103 LTAFFLEA 110

RESULT 7
 ABM69250
 ID ABM69250 standard; protein; 446 AA.

RESULT 8
 ADF06196
 ID ADF06196 standard; protein; 457 AA.
 AC ADF06196;

```

XX 12-FEB-2004 (first entry)
XX Bacterial polypeptide #2309.
DE Proteus mirabilis infection; bacterial infection; antibacterial;
XX Immunostimulant.
XX Proteus mirabilis.
XX OS
XX US6605709-B1.
XX 12-AUG-2003.
XX 05-APR-2000; 2000US-00543681.
XX 09-APR-1999; 99US-0128706P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Breton GL;
XX WPI: 2003-895291/82.
XX N-PSDB; ADF02024.
XX New Proteus mirabilis polypeptides and polynucleotides, useful as
XX reagents for diagnosis of bacterial disease, as components of
XX antibacterial vaccines, as targets for antibacterial drugs, or as
XX biocontrol agents for plants.
XX Disclousure; SEQ ID NO 6481; 870pp; English.
XX The invention relates to new Proteus mirabilis polypeptides and
XX polynucleotides. The invention also relates to antibodies against the
XX polypeptides, methods for producing the polypeptides, a method of
XX generating vaccines for immunising an individual against P. mirabilis, a
XX method for evaluating a compound for the ability to bind a P. mirabilis
XX polypeptide and a method for screening test compounds for anti-bacterial
XX activity. The polypeptides and polynucleotides are useful as molecular
XX targets for diagnosing, preventing and treating pathological conditions
XX resulting from bacterial infection, as reagents for diagnosis of
XX bacterial diseases, as components of antibacterial vaccines, as targets
XX for antibacterial drugs or as bio-control agents for plants. This
XX sequence represents a Proteus mirabilis polypeptide of the invention.
XX
XX Sequence 457 AA;
XX
XX Query Match 19.5%; Score 8; DB 7; Length 457;
XX Best Local Similarity 100.0%; Pred. No. 3.4;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 34 LTAFPLEA 41
XX 110 LTAFPLEA 117
XX
XX RESULT 9
XX ABU40177
XX ID ABU40177 standard; protein; 467 AA.
XX AC ABU40177;
XX 19-JUN-2003 (first entry)
XX Proteins encoded by Prokaryotic essential gene #25704.
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Pseudomonas putida.
XX WO200277183-A2.
XX 03-OCT-2002.

```

```

XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI: 2003-023926/02.
XX N-PSDB; ACA44047.
XX New antisense nucleic acid, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX Claim 25; SEQ ID NO 68101; 1766pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 467 AA;
XX
XX Query Match 19.5%; Score 8; DB 6; Length 467;
XX Best Local Similarity 100.0%; Pred. No. 3.5;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 34 LTAFPLEA 41
XX 92 LTAFPLEA 99
XX
XX RESULT 10
XX ABO63337
XX ID ABO63337 standard; protein; 473 AA.
XX AC ABO63337;
XX 29-JUL-2004 (first entry)

```

```

XX Klebsiella pneumoniae polypeptide seqid 9854.
DE
XX Recombinant expression vector; transcription regulatory element;
KM Klebsiella pneumoniae protein; antibacterial; Vaccine.
XX
OS Klebsiella pneumoniae.
XX
PN US6610836-B1.
XX
PD 26-AUG-2003.
XX
PF 27-JAN-2000; 2000US-00489039.
XX
PR 29-JAN-1999; 99US-0117747P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton GL, Osborne M;
XX
DR WPI; 2003-895346/82.
DR N-PSDB; ACH96888.
XX
PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
PT preparing a vaccine composition against Klebsiella pneumoniae.
XX
PS Disclosure; SEQ ID NO 9854; 932pp; English.
XX
CC The invention describes a new isolated nucleic acid encoding a Klebsiella
CC pneumoniae polypeptide. Also described are: a recombinant expression
CC vector comprising the nucleic acid, operably linked to a transcription
CC regulatory element; and a cell comprising the recombinant expression
CC vector. The nucleic acid is useful for preparing a vaccine composition
CC against Klebsiella pneumoniae. This is the amino acid sequence of a
CC Klebsiella pneumoniae polypeptide of the invention
XX
SQ Sequence 473 AA;
XX
Query Match 19.5%; Score 8; DB 7; Length 473;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 34 LTAFFLEA 41
DB 111 LTAFFLEA 118
XX
RESULT 11
ID ADF04247 standard; protein; 480 AA.
XX
AC ADF04247;
XX
DT 12-FEB-2004 (first entry)
XX
DE Bacterial polypeptide #360.
XX
KM Proteus mirabilis infection; bacterial infection; antibacterial;
XX immunostimulant.
XX
OS Proteus mirabilis.
XX
PN US6605709-B1.
XX
PD 12-AUG-2003.
XX
PF 05-APR-2000; 2000US-00543681.
XX
PR 09-APR-1999; 99US-0128706P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton GL;

```

```

XX WPI; 2003-895291/82.
DR N-PSDB; ADF00075.
XX
PT New Proteus mirabilis polypeptides and polynucleotides, useful as
PT reagents for diagnosis of bacterial disease, as components of
PT antibacterial vaccines, as targets for antibacterial drugs, or as
PT biocontrol agents for plants.
XX
PS Disclosure; SEQ ID NO 4532; 870pp; English.
XX
CC The invention relates to new Proteus mirabilis polypeptides and
CC polynucleotides. The invention also relates to antibodies against the
CC polypeptides, methods for producing the polypeptides, a method of
CC generating vaccines for immunizing an individual against P. mirabilis, a
CC method for evaluating a compound for the ability to bind a P. mirabilis
CC polypeptide and a method for screening test compounds for anti-bacterial
CC activity. The polypeptides and polynucleotides are useful as molecular
CC targets for diagnosing, preventing and treating pathological conditions
CC resulting from bacterial infection, as reagents for diagnosis of
CC bacterial diseases, as components of antibacterial vaccines, as targets
CC for antibacterial drugs or as bio-control agents for plants. This
CC sequence represents a Proteus mirabilis polypeptide of the invention.
XX
SQ Sequence 480 AA;
XX
Query Match 19.5%; Score 8; DB 7; Length 480;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 34 LTAFFLEA 41
DB 139 LTAFFLEA 146
XX
RESULT 12
ID ABO71657 standard; protein; 492 AA.
XX
AC ABO71657;
XX
DT 29-JUL-2004 (first entry)
XX
DE Pseudomonas aeruginosa polypeptide #3832.
XX
KM Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.
XX
OS Pseudomonas aeruginosa.
XX
PN US6551795-B1.
XX
PD 22-APR-2003.
XX
PF 18-FEB-1999; 99US-00252991.
XX
PR 18-FEB-1998; 98US-0074788P.
XX
PR 27-JUL-1998; 98US-0094190P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX
DR WPI; 2003-615309/58.
DR N-PSDB; ABD05228.
XX
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
PS Disclosure; SEQ ID NO 20403; 455pp; English.
XX
CC The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and

```

CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABO67825-
CC ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The
CC sequence data for this patent did not form part of the printed. The
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
CC
CC
SQ Sequence 492 AA;

Query Match 19.5%; Score 8; DB 7; Length 492;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 34 LTAFFLEA 41
Db 107 LTAFFLEA 114

RESULT 13
ADG36882
ID ADG36882 standard; protein; 416 AA.
XX
XX ADG36882;
AC
XX 26-FEB-2004 (first entry)
XX
XX A. thaliana MID1B protein.
XX
XX MID1B; extension-activated calcium permeable channel; plant; gravity;
XX tropism; gravity sensor; photosynthesis.
XX
XX Arabidopsis thaliana.
XX
XX JP2003180367-A.
XX
XX 02-JUL-2003.
XX
XX 19-DEC-2001; 2001JP-00385513.
XX
XX 19-DEC-2001; 2001JP-00385513.
XX
XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX (RIKA) RIKAGAKU KENKYUSHO.
XX
XX WPI; 2003-819302/77.
XX N-PSDB; ADG36881.
XX
XX Novel AtMID1B gene derived from Arabidopsis thaliana encoding extension-
XX activated calcium permeable channel, useful for maintaining gravity
XX tropism in plant.
XX
XX Claim 2; SEQ ID NO 2; 11pp; Japanese.
XX
XX This invention describes a novel gene (AtMID1B) encoding an extension-
XX activated calcium (Ca2+) permeable channel of higher plant, useful for
XX controlling gravity tropism in plants. The gene is useful as gravity
XX sensor in plant. Transgenic plants containing the gene have improved
XX photosynthetic ability.
XX
XX Sequence 416 AA;

Query Match 17.1%; Score 7; DB 7; Length 416;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 23 KKEBLTS 29
Db 243 KKEBLTS 249

RESULT 14
AAB92908
ID AAB92908 standard; protein; 417 AA.
XX
XX AAB92908;
AC
XX 26-JUN-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:11537.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-00116126.
XX
XX 29-JUL-1999; 99JP-00248036.
XX 27-AUG-1999; 99JP-00300253.
XX 11-JAN-2000; 2000JP-00118776.
XX 02-MAY-2000; 2000JP-00183767.
XX 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
XX length cDNAs defined in the specification, and for the detection and/or
XX diagnosis of the abnormality of the proteins encoded by the full-length
XX cDNAs.
XX
XX Claim 8; SEQ ID NO 11537; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-
XX length cDNAs defined in the specification. Where a primer set comprises:
XX (a) an oligo-dT primer and an oligonucleotide complementary to the
XX complementary strand of a polynucleotide which comprises one of the 5602
XX nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in the
XX specification. The primer sets can be used in antisense therapy and in
XX gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
XX represent human amino acid sequences; and AAH13629 to AAH13632 represent
XX oligonucleotides, all of which are used in the exemplification of the
XX present invention
XX
XX Sequence 417 AA;

Query Match 17.1%; Score 7; DB 4; Length 417;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 27 LTSHSQ 33
 |||||
 DB 46 LTSHSQ 52

Search completed: February 1, 2005, 15:20:13
 Job time : 70.7778 secs

RESULT 15

AAV49137 standard; protein; 431 AA.

AAV49137;

17-JAN-2000 (first entry)

Amino acid sequence of choline-binding protein fragment #1.

Truncated surface binding protein; alpha helix; choline binding protein;
 vaccine; invasive bacterial infection; otitis media; sepsis; meningitis;
 lobar pneumonia infection; antibody; immature immune system;
 immunocompromised.

Streptococcus pneumoniae.

MO995126-A2.

14-OCT-1999.

06-APR-1999; 99MO-US007680.

07-APR-1998; 98US-0080878P.

15-MAY-1998; 98US-0085743P.

(MEDI-) MEDIMUNE INC.

Wizemann TM, Koenig S, Johnson LS;

WPI; 1999-601465/51.

N-PSDB; AAZ31077.

New pneumococcal proteins useful as vaccines and for diagnosis of
 pneumococcal infections.

Claim 10; Page 64-65; 98pp; English.

AAV49137-Y49152 are amino acid sequences that are fragments of choline
 binding proteins (CBP). The fragments of the protein are the alpha helix
 forming parts of the CBPs from Streptococcus pneumoniae. The polypeptides
 do not contain the actual choline binding fragment. The polypeptides and
 the nucleotide sequences that encode them (AAZ31077-231092) are used in
 the invention, which relates to polypeptide truncates of a pneumococcal
 surface binding protein containing the highly conserved immunogenic alpha
 helical portion and no choline binding portion. The polypeptides are used
 as immunogens in a bacterial vaccine. The vaccine can be used for
 preventing (immunizing) or treating invasive bacterial (especially
 pneumococcal) infections, especially otitis media (caused by
 S. pneumoniae), sepsis, meningitis and lobar pneumonia infections.
 Antibodies raised against the polypeptide are useful for detection,
 prevention (passive immunity) and treatment of S. pneumoniae infections.
 The vaccines are especially useful in immunocompromised patients, those
 with an immature immune system, or patients with an on going pneumococcal
 infection. The vaccine avoids unnecessary expense and provides broad
 protection against a range of pneumococcal serotypes and it produces an
 improved and enhanced effect in preventing bacterial infections

Sequence 431 AA;

Query Match 17.1%; Score 7; DB 2; Length 431;

Best Local Similarity 100.0%; Pred. No. 39;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 KKEBLTS 29

DB 74 KKEBLTS 80

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using BW model

Run on: February 1, 2005, 14:58:28 / Search time 17.5556 Seconds
(without alignments)
154.882 Million cell updates/sec

Title: SEQ5VAL1854

Perfect score: 41
Sequence: 1 EKMKNMHGFMSILQEHIG.....MKKEELTSHQSLTAFLEA 41Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size: 0

Total number of hits satisfying chosen parameters: 290433

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Issued Patents AA:*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	19.5	411	4	US-09-489-039A-7912
2	8	19.5	457	4	US-09-543-681A-6481
3	8	19.5	473	4	US-09-489-039A-9854
4	8	19.5	480	4	US-09-543-681A-4532
5	8	19.5	492	4	US-09-252-991A-20403
6	7	17.1	431	4	US-09-286-981B-3
7	7	17.1	588	4	US-08-714-741-42
8	7	17.1	605	4	US-08-714-741-46
9	7	17.1	864	4	US-08-714-741-40
10	7	17.1	1231	4	US-08-714-741-41
11	6	14.6	51	4	US-09-513-999C-4343
12	6	14.6	96	4	US-09-270-767-57083
13	6	14.6	97	4	US-09-621-976-5794
14	6	14.6	134	3	US-09-134-001C-5232
15	6	14.6	142	4	US-09-270-767-41841
16	6	14.6	143	4	US-09-198-452A-6
17	6	14.6	162	3	US-09-134-001C-3670
18	6	14.6	226	3	US-09-176-657-3
19	6	14.6	226	4	US-09-421-299-3
20	6	14.6	230	2	US-08-540-118-1
21	6	14.6	230	4	US-09-185-818-1
22	6	14.6	232	4	US-09-328-352-6245
23	6	14.6	254	4	US-09-489-039A-12987
24	6	14.6	263	3	US-08-845-258-18
25	6	14.6	263	3	US-08-990-571-18
26	6	14.6	263	3	US-08-723-142A-18
27	6	14.6	263	4	US-09-528-784A-18

28	6	14.6	263	4	US-09-569-098A-18	Sequence 18, Appl
29	6	14.6	267	3	US-08-845-258-28	Sequence 28, Appl
30	6	14.6	267	3	US-08-990-571-28	Sequence 28, Appl
31	6	14.6	267	3	US-08-723-142A-28	Sequence 28, Appl
32	6	14.6	267	4	US-09-528-784A-28	Sequence 28, Appl
33	6	14.6	267	4	US-09-569-098A-28	Sequence 28, Appl
34	6	14.6	294	3	US-08-845-258-46	Sequence 46, Appl
35	6	14.6	294	3	US-08-990-571-46	Sequence 46, Appl
36	6	14.6	294	4	US-08-723-142A-46	Sequence 46, Appl
37	6	14.6	294	4	US-09-528-784A-46	Sequence 46, Appl
38	6	14.6	294	4	US-09-569-098A-46	Sequence 46, Appl
39	6	14.6	303	3	US-08-845-258-23	Sequence 23, Appl
40	6	14.6	303	3	US-08-990-571-23	Sequence 23, Appl
41	6	14.6	303	3	US-08-723-142A-23	Sequence 23, Appl
42	6	14.6	303	4	US-09-528-784A-23	Sequence 23, Appl
43	6	14.6	303	4	US-09-569-098A-23	Sequence 23, Appl
44	6	14.6	310	3	US-08-845-258-19	Sequence 19, Appl
45	6	14.6	310	3	US-08-990-571-19	Sequence 19, Appl

ALIGNMENTS

```
RESULT 1
US-09-489-039A-7912
; Sequence 7912, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 7912
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7912
```

Query Match 19.5%; Score 8; DB 4; Length 411;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 LTAFLEA 41
|||||
DB 46 LTAFLEA 53

```
RESULT 2
US-09-543-681A-6481
; Sequence 6481, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR APPLICATION NUMBER: 2000-04-05
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6481
; LENGTH: 457
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6481

Query Match 19.5%; Score 8; DB 4; Length 457;
```

Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 LTAFFLEA 41
|||||
Db 110 LTAFFLEA 117

RESULT 3
US-09-489-039A-9854

; Sequence 9854, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; PRIOR FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 9854

; LENGTH: 473

; TYPE: PRT

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-9854

Query Match

Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 LTAFFLEA 41
|||||
Db 111 LTAFFLEA 118

RESULT 4

US-09-543-681A-4532

; Sequence 4532, Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; PRIOR FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1998-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 4532

; LENGTH: 480

; TYPE: PRT

; ORGANISM: Proteus mirabilis

US-09-543-681A-4532

Query Match

Best Local Similarity 19.5%; Score 8; DB 4; Length 480;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 LTAFFLEA 41
|||||
Db 139 LTAFFLEA 146

RESULT 5

US-09-252-991A-20403

; Sequence 20403, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 20403

; LENGTH: 492

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-20403

Query Match

Best Local Similarity 19.5%; Score 8; DB 4; Length 492;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 LTAFFLEA 41
|||||
Db 107 LTAFFLEA 114

RESULT 6

US-09-286-981B-3

; Sequence 3, Application US/09286981B

; Patent No. 6503511

; GENERAL INFORMATION:

; APPLICANT: Witzmann, Theresa M.

; APPLICANT: Johnson, Leslie S

; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines

; FILE REFERENCE: 469201.396

; CURRENT APPLICATION NUMBER: US/09/286,981B

; PRIOR FILING DATE: 1999-04-06

; PRIOR APPLICATION NUMBER: US 60/085,743

; PRIOR FILING DATE: 1998-05-15

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 3

; LENGTH: 431

; TYPE: PRT

; ORGANISM: Streptococcus pneumoniae

US-09-286-981B-3

Query Match

Best Local Similarity 17.1%; Score 7; DB 4; Length 431;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 KKEELTS 29
|||||
Db 74 KKEELTS 80

RESULT 7

US-08-714-741-42

; Sequence 42, Application US/08714741

; Patent No. 6500613

; GENERAL INFORMATION:

; APPLICANT: Briles, David E.

; APPLICANT: McDaniel, Larry S.

; APPLICANT: Swatilo, Edwin

; APPLICANT: Yoether, Janet

; APPLICANT: Crahn, Marilyn J.

; APPLICANT: Hollingshead, Susan

; APPLICANT: Tart, Rebecca

; APPLICANT: Brooks-Walter, Alexis

; TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,

; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,

; NUMBER OF SEQUENCES: 47

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,741
FILING DATE: 16-SEP-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Frommer Esq., William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454312-2460
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 588 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: amino acid
US-08-714-741-42

Query Match 17.1% Score 7; DB 4; Length 588;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 KKEBLTS 29
Db 323 KKEBLTS 329

RESULT 8
US-08-714-741-46
Sequence 46, Application US/08714741
Patent No. 6500613
GENERAL INFORMATION:
APPLICANT: Briles, David E.
APPLICANT: McDaniel, Larry S.
APPLICANT: Swiatlo, Edwin
APPLICANT: Yotter, Janet
APPLICANT: Crain, Marilyn J.
APPLICANT: Hollingshead, Susan
APPLICANT: Tart, Rebecca
APPLICANT: Brooks-Walter, Alexis
TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
TITLE OF INVENTION: PORTIONS AND PRODUCTS
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,741
FILING DATE: 16-SEP-1996
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Frommer Esq., William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454312-2460
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 605 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: amino acid
US-08-714-741-46

Query Match 17.1% Score 7; DB 4; Length 605;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 KKEBLTS 29
Db 106 KKEBLTS 112

RESULT 9
US-08-714-741-40
Sequence 40, Application US/08714741
Patent No. 6500613
GENERAL INFORMATION:
APPLICANT: Briles, David E.
APPLICANT: McDaniel, Larry S.
APPLICANT: Swiatlo, Edwin
APPLICANT: Yotter, Janet
APPLICANT: Crain, Marilyn J.
APPLICANT: Hollingshead, Susan
APPLICANT: Tart, Rebecca
APPLICANT: Brooks-Walter, Alexis
TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
TITLE OF INVENTION: PORTIONS AND PRODUCTS
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,741
FILING DATE: 16-SEP-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Frommer Esq., William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454312-2460
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 864 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: amino acid
US-08-714-741-40

Query Match 17.1% Score 7; DB 4; Length 864;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 KKEELTS 29
 |||||
 DB 296 KKEELTS 302

RESULT 10
 US-08-714-741-41

; Sequence 41, Application US/08714741
 ; Patent No. 6500613

; GENERAL INFORMATION:

; APPLICANT: Briles, David E.
 ; APPLICANT: McDaniel, Larry S.

; APPLICANT: Swiatlo, Edwin
 ; APPLICANT: Yoether, Janet

; APPLICANT: Crain, Marilyn J.
 ; APPLICANT: Hollingshead, Susan

; APPLICANT: Tart, Rebecca
 ; APPLICANT: Brooks-Walter, Alexis

; TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF
 ; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
 ; TITLE OF INVENTION: PORTIONS AND PRODUCTS

; NUMBER OF SEQUENCES: 47
 ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Curtie, Morris & Safford, P.C.
 ; STREET: 530 Fifth Avenue

; CITY: New York
 ; STATE: New York

; COUNTRY: U.S.
 ; ZIP: 10036

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/714,741
 ; FILING DATE: 16-SEP-1996

; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:

; NAME: Frommer Esq., William S.
 ; REGISTRATION NUMBER: 25,506

; REFERENCE/DOCKET NUMBER: 454312-2460
 ; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 840-3333
 ; TELEFAX: (212) 840-0712

; INFORMATION FOR SEQ ID NO: 41:
 ; SEQUENCE CHARACTERISTICS:

; LENGTH: 1231 amino acids
 ; TYPE: amino acid

; STRANDEDNESS: single
 ; TOPOLOGY: linear

; MOLECULE TYPE: amino acid
 ; US-08-714-741-41

Query Match 17.1% Score 7; DB 4; Length 1231;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 KKEELTS 29
 |||||
 DB 125 KKEELTS 131

RESULT 11
 US-09-513-999C-4343

; Sequence 4343, Application US/09513999C
 ; Patent No. 6783961

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.
 ; APPLICANT: Duclert, A.
 ; APPLICANT: Giordano, J.Y.
 ; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

; Patent No. 6783961
 ; FILE REFERENCE: 59, US2, REG

; CURRENT APPLICATION NUMBER: US/09/513, 999C
 ; CURRENT FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/122,487
 ; PRIOR FILING DATE: 1999-02-26

; NUMBER OF SEQ ID NOS: 36681
 ; SOFTWARE: Patent.pm

; SEQ ID NO 4343
 ; LENGTH: 51

; TYPE: PRT
 ; ORGANISM: Homo sapiens

; FEATURE:
 ; NAME/KEY: SIGNAL

; LOCATION: -36...-1
 ; OTHER INFORMATION: score 5.4
 ; OTHER INFORMATION: seq LTCYFGMSAVAS/OL

Query Match 14.6% Score 6; DB 4; Length 51;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 OSOLTA 36
 |||||
 DB 35 OSOLTA 40

RESULT 12
 US-09-270-767-57083

; Sequence 57083, Application US/09270767
 ; Patent No. 6703491

; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 57083
 ; LENGTH: 56

; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster

; US-09-270-767-57083

Query Match 14.6% Score 6; DB 4; Length 56;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 29 SHOSQL 34
 |||||
 DB 10 SHOSQL 15

RESULT 13
 US-09-621-976-5794

; Sequence 5794, Application US/09621976
 ; Patent No. 6639063

; GENERAL INFORMATION:
 ; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Jobert, S.
 ; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
 ; FILE REFERENCE: GENSEP.054PR2

; CURRENT APPLICATION NUMBER: US/09/621,976
 ; CURRENT FILING DATE: 2000-07-21

; NUMBER OF SEQ ID NOS: 19335
 ; SOFTWARE: Patent.pm

SEQ ID NO 5794
 LENGTH: 97
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: SIGNAL
 LOCATION: -53...-1
 US-09-621-976-5794

Query Match 14.6%; Score 6; DB 4; Length 97;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 QSOULTA 36
 |||||
 Db 76 QSOULTA 81

RESULT 14
 US-09-134-001C-5232
 Sequence 5232, Application US/09134001C
 Patent No. 6380370
 GENERAL INFORMATION:
 APPLICANT: Lynn Doucette-Stamm et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: GTC-007
 CURRENT APPLICATION NUMBER: US/09/134,001C
 CURRENT FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,964
 PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,779
 PRIOR FILING DATE: 1997-08-14
 NUMBER OF SEQ ID NOS: 5674
 SEQ ID NO 5232
 LENGTH: 134
 TYPE: PRT
 ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-5232

Query Match 14.6%; Score 6; DB 3; Length 134;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 VMKKEE 26
 |||||
 Db 108 VMKKEE 113

RESULT 15
 US-09-270-767-41841
 Sequence 41841, Application US/09270767
 Patent No. 6703491
 GENERAL INFORMATION:
 APPLICANT: Homburger et al.
 TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 FILE REFERENCE: File Reference: 7326-094
 CURRENT APPLICATION NUMBER: US/09/270,767
 CURRENT FILING DATE: 1999-03-17
 NUMBER OF SEQ ID NOS: 62517
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 41841
 LENGTH: 142
 TYPE: PRT
 ORGANISM: Drosophila melanogaster
 FEATURE:
 OTHER INFORMATION: Xaa means any amino acid
 US-09-270-767-41841

Query Match 14.6%; Score 6; DB 4; Length 142;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 SHOSOL 34
 |||||
 Db 10 SHOSOL 15

Search completed: February 1, 2005, 15:35:43
 Job time : 17.5556 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: February 1, 2005, 15:09:39 ; Search time 59.7778 Seconds
(without alignments)
247.799 Million cell updates/sec

Title: SEQ5VAL1854

Perfect score: 41
Sequence: 1 EKWKMKHMGPFMSILOEHIG.....MKKEELTSHOSQLTAFLLEA 41Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1608061 seqs, 361289386 residues

Word size: 0
Total number of hits satisfying chosen parameters: 1311380

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

```
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	87.8	1149	US-10-128-558-167	Sequence 167, App
2	8	19.5	443	US-10-282-122A-70087	Sequence 70087, A
3	8	19.5	467	US-10-282-122A-68101	Sequence 68101, A
4	7	17.1	336	US-10-299-636-103	Sequence 103, App
5	7	17.1	431	US-10-254-995-3	Sequence 3, Appl
6	7	17.1	496	US-09-748-875-4	Sequence 4, Appl
7	7	17.1	496	US-09-298-523B-4	Sequence 4, Appl
8	7	17.1	589	US-09-748-875-14	Sequence 14, Appl
9	7	17.1	589	US-09-298-523B-14	Sequence 14, Appl
10	7	17.1	589	US-10-299-636-97	Sequence 97, Appl
11	7	17.1	643	US-10-299-636-95	Sequence 95, Appl
12	7	17.1	670	US-09-748-875-63	Sequence 63, Appl
13	7	17.1	670	US-09-298-523B-63	Sequence 63, Appl

14	7	17.1	690	US-09-748-875-61	Sequence 61, Appl
15	7	17.1	690	US-09-298-523B-61	Sequence 61, Appl
16	7	17.1	691	US-09-748-875-1	Sequence 1, Appl
17	7	17.1	691	US-09-298-523B-1	Sequence 1, Appl
18	7	17.1	701	US-09-748-875-62	Sequence 62, Appl
19	7	17.1	701	US-09-298-523B-62	Sequence 62, Appl
20	7	17.1	707	US-09-748-875-2	Sequence 2, Appl
21	7	17.1	707	US-09-298-523B-2	Sequence 2, Appl
22	7	17.1	711	US-09-748-875-3	Sequence 3, Appl
23	7	17.1	711	US-09-298-523B-3	Sequence 3, Appl
24	7	17.1	774	US-10-291-655-459	Sequence 459, App
25	7	17.1	929	US-09-748-875-60	Sequence 60, Appl
26	7	17.1	929	US-09-298-523B-60	Sequence 60, Appl
27	7	17.1	929	US-10-299-636-94	Sequence 94, Appl
28	7	17.1	929	US-10-437-663-136574	Sequence 136574, Appl
29	6	14.6	56	US-09-764-869-1049	Sequence 1049, Ap
30	6	14.6	56	US-09-764-891-5371	Sequence 5371, Ap
31	6	14.6	56	US-10-091-804-1049	Sequence 1049, Ap
32	6	14.6	56	US-10-227-577-1049	Sequence 1049, Ap
33	6	14.6	58	US-10-425-115-216216	Sequence 216216, A
34	6	14.6	58	US-10-425-115-356680	Sequence 356680, A
35	6	14.6	61	US-10-425-115-210289	Sequence 210289, A
36	6	14.6	61	US-10-425-115-299684	Sequence 299684, A
37	6	14.6	73	US-10-425-115-191450	Sequence 191450, A
38	6	14.6	84	US-10-029-386-29893	Sequence 29893, A
39	6	14.6	92	US-10-425-115-189548	Sequence 189548, A
40	6	14.6	95	US-10-424-599-257004	Sequence 257004, A
41	6	14.6	95	US-10-276-774-1710	Sequence 1710, Ap
42	6	14.6	99	US-10-424-599-162535	Sequence 162535, A
43	6	14.6	99	US-10-424-599-204559	Sequence 204559, A
44	6	14.6	106	US-10-767-701-61193	Sequence 61193, A
45	6	14.6	107	US-10-032-201B-109	Sequence 109, App

ALIGNMENTS

```
RESULT 1
US-10-128-558-167
Sequence 167, Application US/10128558
Publication No. US20040219521A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Wang, Zhiwei
APPLICANT: Wang, Gezhi
APPLICANT: Boyle, Bryan J
APPLICANT: Drmanac, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
FILE REFERENCE: 812A
CURRENT APPLICATION NUMBER: US/10128,558
CURRENT FILING DATE: 2002-04-22
PRIOR APPLICATION NUMBER: US 60/339,453
PRIOR FILING DATE: 2001-12-11
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: PCT/US01/03800
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 09/515,126
PRIOR FILING DATE: 2000-02-28
Remaining Prior Application data removed - See File Wrapper or PALM.
```

NUMBER OF SEQ ID NOS: 412
 SOFTWARE: pc FL_genes Version 6.0
 SEQ ID NO 167
 LENGTH: 1149
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-128-558-167

Query Match 87.8%; Score 36; DB 17; Length 1149;
 Best Local Similarity 100.0%; Pred. No. 1.7e-30;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 NHMGPMSTLQEHIGVWKKEBLTSHQSQTAFLEA 41
 Db 844 NHMGPMSTLQEHIGVWKKEBLTSHQSQTAFLEA 879

RESULT 2
 US-10-282-122A-70087
 Sequence 70087, Application US/10282122A
 Publication No. US20040029129A1
 GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
 APPLICANT: Zamudio, Carlos
 APPLICANT: Malone, Cheryl
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Kari
 APPLICANT: Zyskind, Judith
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John
 APPLICANT: Carr, Grant
 APPLICANT: Yamamoto, Robert
 APPLICANT: Forsyth, R.
 APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 FILE REFERENCE: ELITRA.034A
 CURRENT FILING DATE: 2003-02-20
 PRIOR FILING DATE: 2000-03-21
 PRIOR FILING DATE: 2000-03-21
 PRIOR FILING DATE: 2000-03-21
 PRIOR FILING DATE: 2000-05-23
 PRIOR FILING DATE: 2000-05-23
 PRIOR FILING DATE: 2000-05-26
 PRIOR FILING DATE: 2000-05-26
 PRIOR FILING DATE: 2000-09-06
 PRIOR FILING DATE: 2000-09-06
 PRIOR FILING DATE: 2000-09-06
 PRIOR FILING DATE: 2000-09-09
 PRIOR FILING DATE: 2000-10-23
 PRIOR FILING DATE: 2000-10-23
 PRIOR FILING DATE: 2000-11-27
 PRIOR FILING DATE: 2000-11-27
 PRIOR FILING DATE: 2000-12-22
 PRIOR FILING DATE: 2000-12-22
 PRIOR FILING DATE: 2001-02-09
 PRIOR FILING DATE: 2001-02-09
 PRIOR FILING DATE: 2001-02-16
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 78614
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 70087
 LENGTH: 443
 TYPE: PRT
 ORGANISM: Pseudomonas syringae
 US-10-282-122A-70087

Query Match 19.5%; Score 8; DB 15; Length 443;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 34 LTAFLEA 41
 |||||||

Db 67 LTAFLEA 74

RESULT 3
 US-10-282-122A-68101
 Sequence 68101, Application US/10282122A
 Publication No. US20040029129A1
 GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
 APPLICANT: Zamudio, Carlos
 APPLICANT: Malone, Cheryl
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Kari
 APPLICANT: Zyskind, Judith
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John
 APPLICANT: Carr, Grant
 APPLICANT: Yamamoto, Robert
 APPLICANT: Forsyth, R.
 APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 FILE REFERENCE: ELITRA.034A
 CURRENT FILING DATE: 2003-02-20
 PRIOR FILING DATE: 2000-03-21
 PRIOR FILING DATE: 2000-03-21
 PRIOR FILING DATE: 2000-03-21
 PRIOR FILING DATE: 2000-05-23
 PRIOR FILING DATE: 2000-05-23
 PRIOR FILING DATE: 2000-05-26
 PRIOR FILING DATE: 2000-05-26
 PRIOR FILING DATE: 2000-09-06
 PRIOR FILING DATE: 2000-09-06
 PRIOR FILING DATE: 2000-09-06
 PRIOR FILING DATE: 2000-09-09
 PRIOR FILING DATE: 2000-10-23
 PRIOR FILING DATE: 2000-10-23
 PRIOR FILING DATE: 2000-11-27
 PRIOR FILING DATE: 2000-11-27
 PRIOR FILING DATE: 2000-12-22
 PRIOR FILING DATE: 2000-12-22
 PRIOR FILING DATE: 2001-02-09
 PRIOR FILING DATE: 2001-02-09
 PRIOR FILING DATE: 2001-02-16
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 78614
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 68101
 LENGTH: 467
 TYPE: PRT
 ORGANISM: Pseudomonas putida
 US-10-282-122A-68101

Query Match 19.5%; Score 8; DB 15; Length 467;
 Best Local Similarity 100.0%; Pred. No. 3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 34 LTAFLEA 41
 |||||||
 Db 92 LTAFLEA 99

RESULT 4
 US-10-299-636-103
 Sequence 103, Application US/10299636
 Publication No. US20040077847A1
 GENERAL INFORMATION:

APPLICANT: Briles, David E
 APPLICANT: McDaniel, Larry S
 APPLICANT: Swatlow, Edwin
 APPLICANT: Yohner, Janet
 APPLICANT: Crahn, Marilyn J
 APPLICANT: Hollingshead, Susan

APPLICANT: Tart, Rebecca
APPLICANT: Brooks-Walter, Alexis
TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEINS AND USES THEREOF
FILE REFERENCE: 57909/361
CURRENT APPLICATION NUMBER: US/10/299,636
CURRENT FILING DATE: 2002-11-19
PRIOR APPLICATION NUMBER: 08/714,741
PRIOR FILING DATE: 1996-09-16
PRIOR APPLICATION NUMBER: 08/529,055
PRIOR FILING DATE: 1995-09-15
NUMBER OF SEQ ID NOS: 111
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 103
LENGTH: 336
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-10-299-636-103

Query Match 17.1%; Score 7; DB 15; Length 336;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 23 KKEBLTS 29
Db 56 KKEBLTS 62

RESULT 5
US-10-254-995-3
Sequence 3, Application US/10254995
Publication No. US20030138447A1
GENERAL INFORMATION:
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/10/254,995
CURRENT FILING DATE: 2002-09-25
PRIOR APPLICATION NUMBER: US/09/286,981
PRIOR FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 3
LENGTH: 431
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-10-254-995-3

Query Match 17.1%; Score 7; DB 14; Length 431;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 23 KKEBLTS 29
Db 74 KKEBLTS 80

RESULT 6
US-09-748-875-4
Sequence 4, Application US/09748875
Publication No. US20010016200A1
GENERAL INFORMATION:
APPLICANT: BRILES et al.
TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
FILE REFERENCE: 454312-3140
CURRENT APPLICATION NUMBER: US/09/748,875
CURRENT FILING DATE: 2000-12-26
PRIOR APPLICATION NUMBER: 09/298,523
PRIOR FILING DATE: 1999-04-23

NUMBER OF SEQ ID NOS: 78
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 4
LENGTH: 496
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-748-875-4

Query Match 17.1%; Score 7; DB 9; Length 496;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 23 KKEBLTS 29
Db 120 KKEBLTS 126

RESULT 7
US-09-298-523B-4
Sequence 4, Application US/09298523B
Publication No. US20030059438A1
GENERAL INFORMATION:
APPLICANT: BRILES et al.
TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
FILE REFERENCE: 454312-3140
CURRENT APPLICATION NUMBER: US/09/298,523B
CURRENT FILING DATE: 1999-04-23
NUMBER OF SEQ ID NOS: 78
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 4
LENGTH: 496
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-298-523B-4

Query Match 17.1%; Score 7; DB 10; Length 496;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 23 KKEBLTS 29
Db 120 KKEBLTS 126

RESULT 8
US-09-748-875-14
Sequence 14, Application US/09748875
Publication No. US20010016200A1
GENERAL INFORMATION:
APPLICANT: BRILES et al.
TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
FILE REFERENCE: 454312-3140
CURRENT APPLICATION NUMBER: US/09/748,875
CURRENT FILING DATE: 2000-12-26
PRIOR APPLICATION NUMBER: 09/298,523
PRIOR FILING DATE: 1999-04-23
NUMBER OF SEQ ID NOS: 78
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 14
LENGTH: 589
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-748-875-14

Query Match 17.1%; Score 7; DB 9; Length 589;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 23 KKEBLTS 29
Db 324 KKEBLTS 330

RESULT 9
US-09-298-523B-14
; Sequence 14, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 589
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-14

Query Match 17.1%; Score 7; DB 10; Length 589;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 KKEELTS 29
Db 324 KKEELTS 330

RESULT 10
US-10-299-636-97
; Sequence 97, Application US/10299636
; Publication No. US20040077847A1
; GENERAL INFORMATION:
; APPLICANT: Briles, David E
; APPLICANT: McDaniel, Larry S
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yoether, Janet
; APPLICANT: Crain, Marilyn J
; APPLICANT: Hollingshead, Susan
; APPLICANT: Tart, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEINS AND USES THEREOF
; FILE REFERENCE: 57909/361
; CURRENT APPLICATION NUMBER: US/10/299,636
; CURRENT FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: 08/714,741
; PRIOR FILING DATE: 1996-09-16
; PRIOR APPLICATION NUMBER: 08/529,055
; PRIOR FILING DATE: 1995-09-15
; NUMBER OF SEQ ID NOS: 111
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 97
; LENGTH: 589
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-299-636-97

Query Match 17.1%; Score 7; DB 15; Length 589;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 KKEELTS 29
Db 324 KKEELTS 330

RESULT 11
US-10-299-636-95
; Sequence 95, Application US/10299636
; Publication No. US20040077847A1
; GENERAL INFORMATION:

; APPLICANT: Briles, David E
; APPLICANT: McDaniel, Larry S
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yoether, Janet
; APPLICANT: Crain, Marilyn J
; APPLICANT: Hollingshead, Susan
; APPLICANT: Tart, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEINS AND USES THEREOF
; FILE REFERENCE: 57909/361
; CURRENT APPLICATION NUMBER: US/10/299,636
; CURRENT FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: 08/714,741
; PRIOR FILING DATE: 1996-09-16
; PRIOR APPLICATION NUMBER: 08/529,055
; PRIOR FILING DATE: 1995-09-15
; NUMBER OF SEQ ID NOS: 111
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 95
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-299-636-95

Query Match 17.1%; Score 7; DB 15; Length 643;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 KKEELTS 29
Db 76 KKEELTS 82

RESULT 12
US-09-748-875-63
; Sequence 63, Application US/09748875
; Publication No. US20010016200A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/748,875
; CURRENT FILING DATE: 2000-12-26
; PRIOR APPLICATION NUMBER: 09/298,523
; PRIOR FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 670
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-748-875-63

Query Match 17.1%; Score 7; DB 9; Length 670;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 KKEELTS 29
Db 125 KKEELTS 131

RESULT 13
US-09-298-523B-63
; Sequence 63, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B

; CURRENT FILING DATE: 1999-04-23
 ; NUMBER OF SEQ ID NOS: 78
 ; SOFTWARE: Patentln Ver. 2.1
 ; SEQ ID NO 63
 ; LENGTH: 670
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 US-09-298-523B-63

Query Match 17.1%; Score 7; DB 10; Length 670;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 KKEELTS 29
 |||||
 Db 125 KKEELTS 131

RESULT 14
 US-09-748-875-61
 ; Sequence 61, Application US/09748875
 ; Publication No. US20010016200A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BRILES et al.
 ; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
 ; FILE REFERENCE: 454312-3140
 ; CURRENT APPLICATION NUMBER: US/09/748,875
 ; CURRENT FILING DATE: 2000-12-26
 ; PRIOR APPLICATION NUMBER: 09/298,523
 ; PRIOR FILING DATE: 1999-04-23
 ; NUMBER OF SEQ ID NOS: 78
 ; SOFTWARE: Patentln Ver. 2.1
 ; SEQ ID NO 61
 ; LENGTH: 690
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 US-09-748-875-61

Query Match 17.1%; Score 7; DB 9; Length 690;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 KKEELTS 29
 |||||
 Db 361 KKEELTS 367

RESULT 15
 US-09-298-523B-61
 ; Sequence 61, Application US/09298523B
 ; Publication No. US20030059438A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BRILES et al.
 ; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
 ; FILE REFERENCE: 454312-3140
 ; CURRENT APPLICATION NUMBER: US/09/298,523B
 ; CURRENT FILING DATE: 1999-04-23
 ; NUMBER OF SEQ ID NOS: 78
 ; SOFTWARE: Patentln Ver. 2.1
 ; SEQ ID NO 61
 ; LENGTH: 690
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 US-09-298-523B-61

Query Match 17.1%; Score 7; DB 10; Length 690;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 KKEELTS 29
 |||||

Db 361 KKEELTS 367

Search completed: February 1, 2005, 15:44:48
 Job time : 59.7778 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using BW model

Run on: February 1, 2005, 14:55:01 ; Search time 13.5556 Seconds
(without alignments)
291.016 Million cell updates/sec

Title: SEQ5VAL1854
Perfect score: 41
Sequence: 1 EKXWKNNMGPFMSILOEHG.....MKKEELTSHQSOLTAFLLEA 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 279530

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :
1: PIR_79:*
2: pir1:*
3: pir2:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	19.5	467	2	AC0546
2	8	19.5	488	2	T47273
3	7	17.1	276	2	T08859
4	7	17.1	417	2	D84556
5	7	17.1	537	2	A38170
6	7	17.1	960	2	S44812
7	7	17.1	1196	2	H66389
8	7	17.1	1478	2	S78131
9	6	14.6	40	2	T06840
10	6	14.6	69	2	S04666
11	6	14.6	75	2	S75054
12	6	14.6	107	1	A26622
13	6	14.6	134	2	S75307
14	6	14.6	137	2	T46439
15	6	14.6	143	2	B86624
16	6	14.6	143	2	G72000
17	6	14.6	160	2	T44921
18	6	14.6	189	2	H96520
19	6	14.6	219	1	Z0ZQMR
20	6	14.6	223	1	B82911
21	6	14.6	228	2	AC3204
22	6	14.6	237	2	G70066
23	6	14.6	241	2	G97416
24	6	14.6	241	2	AT2634
25	6	14.6	274	1	P40362
26	6	14.6	276	2	S27641
27	6	14.6	280	2	D86193
28	6	14.6	351	2	T03153
29	6	14.6	358	2	J00596

30	6	14.6	372	2	S76427	hypothetical probe
31	6	14.6	377	2	AD1937	permease protein o
32	6	14.6	383	2	E96972	riboflavin biosynt
33	6	14.6	398	2	S13269	translation initia
34	6	14.6	402	2	S30278	translation initia
35	6	14.6	406	1	FTMS4A	translation initia
36	6	14.6	406	2	S33681	translation initia
37	6	14.6	407	2	S00985	conserved hypothet
38	6	14.6	428	2	B81531	CR850 hypothetical
39	6	14.6	432	2	D72008	CR850 hypothetical
40	6	14.6	432	2	B86616	hypothetical prote
41	6	14.6	433	2	B90419	hypothetical prote
42	6	14.6	436	2	T30911	xylinase (EC 3.2.1
43	6	14.6	441	2	A83370	probable MFS trans
44	6	14.6	443	2	F83759	cytochrome d (bd-t
45	6	14.6	450	2	T04265	probable kasein ki

ALIGNMENTS

RESULT 1
AC0546
Probable terminal oxidase chain I [imported] - Salmonella enterica subsp. enterica serov
C:Species: Salmonella enterica subsp. enterica serovar Typh
A:Note: this species has also been called Salmonella typh
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AC0546
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Garra, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AC0546
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-467 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08815.1; PID:G16501629; GSPDB:GN00176
C:Gene: STY0392
C:Superfamily: cytochrome d complex terminal oxidase chain I
Query Match 19.5% Score 8; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 0.72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 34 LTAFLEA 41
DB 102 LTAFLEA 109
RESULT 2
T47273
Cyanide insensitive terminal oxidase chain c10A [imported] - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C:Accession: T47273; G83155
R:Cunningham, L.; Pitt, M.; Williams, H.D.
Mol. Microbiol. 24, 579-591, 1997
A:Title: The c10AB genes from Pseudomonas aeruginosa code for a novel cyanide-insensitiv
A:Reference number: Z24440; MUID:97323403; PMID:9179851
A:Accession: T47273
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-488 <CUN>
A:Cross-references: UNIPROT:O07440; EMBL:Y10528; NID:G2208963; PIDN:CAA71555.1; PID:G220
A:Experimental source: strain PA01, substrain PA06049
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidig, K.; Lim,
J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000

A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic path
 A:Reference number: AB2950; MUID:20437337; PMID:10984043
 A:Accession: G83155
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-488 <STO>
 A:Cross-references: GB:AE004810; GB:AE004091; NID:g9950106; PIDN:AA07317.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Gene: ctoA
 C:Superfamily: cytochrome d complex terminal oxidase chain I

Query Match 19.1%; Score 8; DB 2; Length 488;
 Best Local Similarity 100.0%; Pred. No. 0.75;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 LTAFFLEA 41
 DB 103 LTAFFLEA 110

RESULT 3
 T08859
 hypothetical protein A_TM017A05.4 - *Arabidopsis thaliana*
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)
 C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 23-Jul-1999
 C:Accession: T08859
 R:Waterston, R.
 submitted to the EMBL Data Library, October 1997
 A:Reference number: Z16500
 A:Accession: T08859
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-276 <MAT>
 A:Cross-references: EMBL:AF024504; NID:g2435510; PID:g2435512
 C:Genetics:
 A:Map position: 4
 A:Introns: 137/3
 A:Note: A_TM017A05.4

Query Match 17.1%; Score 7; DB 2; Length 276;
 Best Local Similarity 100.0%; Pred. No. 5.4;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 KKEBLTS 29
 DB 243 KKEBLTS 249

RESULT 4
 D84556
 hypothetical protein At2g17780 [imported] - *Arabidopsis thaliana*
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)
 C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: D84556
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Conlin, L.A.; Shen, M.; VanKen, S.E.; Unayam, L.; Tallon, L.;
 euss, D.; Mierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A>Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
 A:Reference number: AB4420; MUID:20083487; PMID:10617157
 A:Accession: D84556
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-417 <STO>
 A:Cross-references: GB:AE002093; NID:g6598810; PIDN:AB80787.2; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g17780
 A:Map position: 2

Query Match 17.1%; Score 7; DB 2; Length 417;
 Best Local Similarity 100.0%; Pred. No. 7.8;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 KKEBLTS 29
 DB 243 KKEBLTS 249

RESULT 5
 A38170
 cytochrome d complex terminal oxidase (EC 1.10.3.-) chain I - *Azotobacter vinelandii*
 C:Species: *Azotobacter vinelandii*
 C>Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 09-Jul-2004
 C:Accession: A38170; A41578
 R:Moshiri, F.; Smith, E.G.; Taormino, J.P.; Maier, R.J.
 J. Bacteriol. 173, 6230-6241, 1991
 A>Title: Cloning, characterization, and expression in *Escherichia coli* of the genes enco
 A:Reference number: A38170; MUID:92011387; PMID:1655703
 A:Accession: A38170
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-537 <MOS>
 A:Cross-references: UNIPROT:Q09049; GB:S57066; NID:g236633; PIDN:AB19986.1; PID:g236634
 R:Moshiri, F.; Smith, E.G.; Taormino, J.P.; Maier, R.J.
 J. Biol. Chem. 266, 23169-23174, 1991
 A>Title: Transcriptional regulation of cytochrome d in nitrogen-fixing *Azotobacter vine*
 A:Reference number: A41578; MUID:92078187; PMID:1660468
 A:Accession: A41578
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-13 <MO2>
 A:Cross-references: GB:M77787
 A:Note: the authors translated the codon CAG for residue 13 as Gly
 C:Superfamily: cytochrome d complex terminal oxidase chain I
 C:Keywords: oxidoreductase

Query Match 17.1%; Score 7; DB 2; Length 537;
 Best Local Similarity 100.0%; Pred. No. 9.8;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 LTAFFLE 40.
 DB 103 LTAFFLE 109

RESULT 6
 S44812
 F4489.6 protein - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 04-Mar-2000
 C:Accession: S44812
 R:Antonacci-Fulton, L.
 submitted to the EMBL Data Library, September 1993
 A:Description: Sequence of the *C. elegans* cosmid F4489.
 A:Reference number: S44807
 A:Accession: S44812
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-960 <ANT>
 A:Cross-references: EMBL:L23648; NID:g388585; PID:g388591
 C:Genetics:
 A:Introns: 33/3; 66/3; 261/2; 299/3; 479/2; 615/3; 750/3; 903/2; 925/3
 C:Superfamily: *Caenorhabditis elegans* F4489.6 protein

Query Match 17.1%; Score 7; DB 2; Length 960;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 ELTSHOS 32
 DB 478 ELTSHOS 484

RESULT 7
 H86389

hypothetical protein F28823.2 - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
 C/Accession: F06840
 R/Phylogeny: A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Hansen, N.F.; Hughes, B.; Hultine, L.
 Nature 408, 816-820, 2000
 A/Authors: Hunter, J.L.; Jenkin, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luro, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A/Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A/Reference number: A66141; MUID:21016719; PMID:11130712
 A/Accession: H66389
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-1196 <STO>
 A/Cross-references: UNIPROT:Q9C678; GB:AE005172; NID:g11079511; PIDN:AG29222.1; GSPDB:G
 C/Genetics:
 A/Map position: 1

Query Match 17.1%; Score 7; DB 2; Length 1196;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 MKKELT 28
 |||||
 Db 349 MKKELT 355

RESULT 8
 S78131
 DNA-directed RNA polymerase (EC 2.7.7.6) chain beta - Recclinomonas americana (ATCC 50394)
 C/Species: mitochondrion Recclinomonas americana
 A/Variety: ATCC 50394
 C/Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 09-Jul-2004
 C/Accession: S78131
 R/Liang, B.F.; Burger, G.; O'Kelly, C.J.; Cedergren, R.; Golding, G.B.; Lemieux, C.; Sank
 Nature 387, 493-497, 1997
 A/Title: An ancestral mitochondrial DNA resembling a eubacterial genome in miniature.
 A/Reference number: S78127; MUID:97311393; PMID:9168110
 A/Accession: S78131
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-1478 <LAN>
 A/Cross-references: UNIPROT:Q21337; EMBL:AF007261; NID:g2258325; PIDN:AAD11864.1; PID:g2
 A/Experimental source: ATCC 50394
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1997
 C/Genetics:
 A/Gene: rpoB
 A/Genome: mitochondrion
 A/Suprafamily: DNA-directed RNA polymerase beta chain
 C/KeyWords: mitochondrion; nucleotidyltransferase; transcription

Query Match 17.1%; Score 7; DB 2; Length 1478;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 KKEELT 29
 |||||
 Db 426 KKEELT 432

RESULT 9
 T06840
 photosystem I chain IX - Cyanophora paradoxa cyanelle
 C/Species: cyanelle Cyanophora paradoxa
 C/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004
 C/Accession: T06840
 R/Strehlitz, V.L.; Michalowski, C.B.; Luffelhardt, W.; Bohner, H.J.; Bryant, D.A.
 submitted to the EMBL Data Library, July 1995

A/Description: Nucleotide sequence of the cyanelle genome from Cyanophora paradoxa.
 A/Reference number: Z15840
 A/Accession: T06840
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-40 <STI>
 A/Cross-references: UNIPROT:P48117; EMBL:U30921; NID:g1016083; PIDN:AAA81183.1; PID:g101
 A/Experimental source: strain Pringsheim LB555
 C/Genetics:
 A/Gene: psal
 A/Genome: cyanelle
 C/Suprafamily: photosystem I protein psal
 C/KeyWords: cyanelle; photosynthesis; photosystem I; thylakoid

Query Match 14.6%; Score 6; DB 2; Length 40;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 LTAFL 39
 |||||
 Db 16 LTAFL 21

RESULT 10
 S04666
 hypothetical protein 1 - Rhodospseudomonas blautica (fragment)
 C/Species: Rhodospseudomonas blautica
 C/Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
 C/Accession: S04666
 R/Tybuliewicz, V.L.; Falk, G.; Walker, J.E.
 J. Mol. Biol. 179, 185-214, 1984
 A/Title: Rhodospseudomonas blautica atp operon. Nucleotide sequence and transcription.
 A/Reference number: S04666; MUID:85058188; PMID:6209404
 A/Accession: S04666
 A/Status: not compared with conceptual translation
 A/Status: not compared with conceptual translation
 A/Molecule type: DNA
 A/Residues: 1-69 <TYB>
 A/Cross-references: UNIPROT:P05443

Query Match 14.6%; Score 6; DB 2; Length 69;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 EELTSH 30
 |||||
 Db 3 EELTSH 8

RESULT 11
 S75054
 hypothetical protein ssl127 - Synechocystis sp. (strain PCC 6803)
 C/Species: Synechocystis sp.
 A/Variety: PCC 6803
 C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
 C/Accession: S75054
 R.Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.;
 O.K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
 S.
 A/Reference number: S74322; MUID:97061201; PMID:8905231
 A/Accession: S75054
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-75 <KAN>
 A/Cross-references: UNIPROT:P73857; EMBL:D90910; GB:AB001339; NID:g1652956; PIDN:BA11791
 C/Genetics:
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 A/Start codon: GTG

Query Match 14.6%; Score 6; DB 2; Length 75;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 KKEBLT 28
|||||
Db 67 KKEBLT 72

RESULT 12

A26622
thioredoxin - Chromatium vinosum
C:Species: Chromatium vinosum
C:Date: 31-Mar-1988 #sequence_revision 26-May-1994 #text_change 16-Aug-2004
C:Accession: A26622
R:Johnson, R.S.; Bleemann, K.
Biochemistry 26, 1209-1214, 1987
A:Title: The primary structure of thioredoxin from Chromatium vinosum determined by high
A:Reference number: A26622; PMID:87185419; PMID:3567166
A:Accession: A26622
A:Molecule type: protein
A:Residues: 1-107 <OOH>
A:Cross-references: UNIPROT:P09857
A>Note: unidentified residues are Ile or Leu
C:Superfamily: Thioredoxin; thioredoxin homology
C:Keywords: heat-stable protein; redox-active disulfide
F:10-93/Domains: thioredoxin homology <THR>
F:32-35/Disulfide bonds: redox-active #status predicted

Query Match

Best Local Similarity 14.6%; Score 6; DB 1; Length 107;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 SOLTAF 37
|||||
Db 97 SOLTAF 102

RESULT 13

S75307
ferredoxin [2Fe-2S] s111584 [similarity] - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C:Accession: S75307
R:Kaneko, T.; Saito, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
sp.
A:Reference number: S74322; PMID:97061201; PMID:8905231
A:Accession: S75307
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-134 <KAN>
A:Cross-references: UNIPROT:P71395; EMBL:D90904; GB:AB001339; NID:G1652225; PIDN:BAA1722
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: ferredoxin [2Fe-2S], Clostridium type
C:Keywords: 2Fe-2S; electron transfer; iron-sulfur protein; metalloprotein
F:32,40,80,84/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match

Best Local Similarity 14.6%; Score 6; DB 2; Length 134;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 ILQEH 19
|||||
Db 110 ILQEH 115

RESULT 14

T46439
hypothetical protein DKFZp434M0326.1 - human
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C:Accession: T46439

R:Ansoorge, W.; Winkler, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23028
A:Accession: T46439

A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-137 <AAA>
A:Cross-references: UNIPROT:Q9NSX2; EMBL:AL137681
A:Experimental source: adult testis; clone DKFZp434M0326
C:Genetics:
A>Note: DKFZp434M0326.1

Query Match

Best Local Similarity 14.6%; Score 6; DB 2; Length 137;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 KKEBLT 28
|||||
Db 22 KKEBLT 27

RESULT 15

B86624
HH transcription regulator [imported] - Chlamydomonas pneumoniae (strain J138)
C:Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: B86624
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.
A:Reference number: A86491; PMID:20330349; PMID:10871362
A:Accession: B86624
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-143 <STO>
A:Cross-references: UNIPROT:Q9Z6T0; GB:BA000008; NID:98979442; PIDN:BAA99276.1; GSPDB:GN
A:Experimental source: strain J138
C:Genetics:
A:Gene: yfGA

Query Match

Best Local Similarity 14.6%; Score 6; DB 2; Length 143;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 SILOEH 18
|||||
Db 75 SILOEH 80

Search completed: February 1, 2005, 15:33:02
Job time: 13.5556 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 1, 2005, 14:54:06 ; Search time 70.5556 Seconds
(without alignments)
334.352 Million cell updates/sec

Title: SEQ5VAL1854
Perfect score: 41
Sequence: 1 EKWNKNHMGPFMSILQEHIG.....MKKEELTSHOSQLTAFLEA 41

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1825181 seqs, 575374646 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1810864

Minimum DB seq length: 21

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Uniprot_02:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	958	1 BP28_MACFA	Q9GM44 macaca fasc
2	41	100.0	2144	1 BP28_HUMAN	Q9S583 homo sapien
3	26	63.4	349	2 Q8VCK1	Q8VCK1 mus musculu
4	26	63.4	408	2 Q8CCT5	Q8CCT5 mus musculu
5	20	48.8	349	2 Q96E55	Q96E55 homo sapien
6	20	48.8	897	2 Q8N7L7	Q8N7L7 homo sapien
7	20	48.8	1106	2 Q6P197	Q6P197 homo sapien
8	20	48.8	1106	2 AAH65205	AAH65205 homo sapi
9	8	19.5	163	2 Q8B1C5	Q8B1C5 mus musculu
10	8	19.5	354	2 Q8B1J2	Q8B1J2 mus musculu
11	8	19.5	445	2 Q7N5B2	Q7N5B2 photorhabdu
12	8	19.5	448	2 Q7NRX0	Q7NRX0 chromobacte
13	8	19.5	453	2 Q98DM1	Q98DM1 rhizobium 1
14	8	19.5	454	2 Q87H28	Q87H28 vibrio para
15	8	19.5	462	2 Q7MB10	Q7MB10 bordetella
16	8	19.5	462	2 Q7WN02	Q7WN02 bordetella
17	8	19.5	462	2 Q8XQ13	Q8XQ13 ralsstonia s
18	8	19.5	465	2 Q86080	Q86080 rhododactyl
19	8	19.5	465	2 Q89NV6	Q89NV6 bradyrhizob
20	8	19.5	465	2 Q8P701	Q8P701 xanthomonas
21	8	19.5	467	2 Q8Z912	Q8Z912 baltomella
22	8	19.5	467	2 Q8ZRG3	Q8ZRG3 baltomella
23	8	19.5	468	2 Q6N0G9	Q6N0G9 rhodopseudo
24	8	19.5	468	2 Q7J3H5	Q7J3H5 wolbachia p
25	8	19.5	468	2 CAE30233	CAE30233 rhodopseu
26	8	19.5	468	2 AAS14431	AAS14431 wolbachia
27	8	19.5	469	2 Q8P682	Q8P682 xanthomonas
28	8	19.5	471	2 Q8P4L6	Q8P4L6 xanthomonas
29	8	19.5	478	2 Q8BE17	Q8BE17 pseudomonas
30	8	19.5	479	2 Q87WA6	Q87WA6 pseudomonas
31	8	19.5	487	2 Q8P1B5	Q8P1B5 xanthomonas

32	8	19.5	488	2	Q07440	Q07440 pseudomonas
33	8	19.5	488	2	Q7DC90	Q7DC90 pseudomonas
34	8	19.5	493	2	Q92S67	Q92S67 rhizobium m
35	7	17.1	258	2	Q8BE22	Q8BE22 pseudomonas
36	7	17.1	296	2	Q8S312	Q8S312 phytophthor
37	7	17.1	405	2	Q8GFD0	Q8GFD0 clostridium
38	7	17.1	417	2	Q9NV74	Q9NV74 homo sapien
39	7	17.1	460	2	Q7Z9N1	Q7Z9N1 trichoderma
40	7	17.1	464	2	Q8NKT5	Q8NKT5 acidianus a
41	7	17.1	482	2	Q8XQ15	Q8XQ15 ralsstonia s
42	7	17.1	525	2	Q8GZT5	Q8GZT5 emiliania h
43	7	17.1	525	2	Q8BUT6	Q8BUT6 lactobacill
44	7	17.1	537	1	CYDA_AZOV1	Q09049 azotobacter
45	7	17.1	545	2	Q6BU55	Q6BU55 debaryomyce

ALIGNMENTS

```

RESULT 1
BP28_MACFA          STANDARD;          PRT;          958 AA.
ID   Q9GM44
AC   Q9GM44
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DE   Protein BAP28 (Onpa-17571) (Fragment).
GN   Name=BAP28;
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Texao K.,
RA   Suzuki Y., Sugano S., Hashimoto K.;
RT   "Isolation of full-length cDNA clones from macaque brain cDNA
RT   libraries.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC   -1- SIMILARITY: Belongs to the BAP28 family.
CC   -1- SIMILARITY: Contains 1 HEAT repeat.
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@sib.ch).
CC   -----
CC   EMBL; AB049842; BAB16728.1; ALT_INIT.
CC   InterPro; IPR008938; ARM.
CC   InterPro; IPR000357; HEAT.
CC   PROSITE; PS50077; HEAT_REPEAT; FALSE_NEG.
FT   NON_TER
FT   REPEAT
FT   REPEAT
SQ   SEQUENCE 958 AA; 108644 MW; 3DBD95C3623CFB31 CRC64;
Query Match          100.0%; Score 41; DB 1; Length 958;
Best Local Similarity 100.0%; Pred. No. 9.8e-36;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY   1 EKWNKNHMGPFMSILQEHIGVWKKKEELTSHOSQLTAFLEA 41
DB   648 EKWNKNHMGPFMSILQEHIGVWKKKEELTSHOSQLTAFLEA 688
RESULT 2
BP28_HUMAN          STANDARD;          PRT;          2144 AA.
ID   Q9H583
AC   Q9H583; Q9NW23;

```

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Protein BAP28.
 GN Name=BAP28.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANTS SER-1694; ALA-1854; ASP-1967 AND
 RP GLY-2017.
 RA Bouqueret L., Chumakov I., Barry C., Cohen-Akenine A.;
 RA "A novel BAP28 gene and protein."
 RA Patent number WO0100669, 04-JAN-2001.
 RN [2]
 RP SEQUENCE OF 1534-2144 FROM N.A.
 RA Cobley V.;
 RA Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1777-2144 FROM N.A.
 RX PubMed=14702039; DOI=10.1038/ng1285;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Ohtsuka M., Nishitani T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J., I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
 RA Nagahara K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
 RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
 RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
 RA Omura Y., Abe K., Kamihara K., Katsura N., Sato K., Tanikawa M.,
 RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
 RA Fujimori T., Tanai H., Kimata M., Watanabe M., Hirao K., Chiba Y.,
 RA Ishida S., Oono Y., Takiguchi S., Watanabe S., Yosida M., Houchi T.,
 RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
 RA Nomura Y., Togiya K., Komai F., Hara R., Takeuchi K., Arita M.,
 RA Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotaka S.,
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata H., Sano S.,
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Omori Y.,
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
 RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
 RA "Complete sequencing and characterization of 21,243 full-length human
 RT cDNAs."
 RN Nat. Genet. 36:40-45(2004).
 CC -1- SIMILARITY: Belongs to the BAP28 family.
 CC -1- SIMILARITY: Contains 1 HEAT repeat.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL; AK067150; CAC26776.1; -;
 DR EMBL; AL136105; CAC15948.1; -;
 DR EMBL; AK001221; BAA91564.1; ALT_INIT.
 DR SWISS-2DPAGE; Q9H583; HUMAN.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR000357; HEAT.
 DR PROSITE; PSS0077; HEAT_REPEAT; FALSE_NEG.
 KW Polymorphism.
 FT REPEAT 2106 2142 HEAT.

FT VARIANT 1694 1694 N -> S.
 FT FTID=VAR_010939.
 FT VARIANT 1854 1854 V -> A.
 FT FTID=VAR_010940.
 FT VARIANT 1967 1967 N -> D.
 FT FTID=VAR_010941.
 FT VARIANT 2017 2017 E -> G.
 FT FTID=VAR_010942.
 SQ SEQUENCE 2144 AA; 242355 MW; D668165E78D9C9B7 CRC64;
 Query Match 100.0%; Score 41; DB 1; Length 2144;
 Best Local Similarity 100.0%; Pred. No. 2e-35;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EKNKKNHMGPPMSILQEHIGVKKKEELTSHQSOLTAFFLEA 41
 DB 1834 EKNKKNHMGPPMSILQEHIGVKKKEELTSHQSOLTAFFLEA 1874
 RESULT 3
 ID Q8VCK1 PRELIMINARY; PRT; 349 AA.
 AC Q8VCK1;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE BC019693 protein.
 GN Name=BC019693;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=mx FVB/N;
 RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shermen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diachenko L., Marusha K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udell T.B., Tohiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=mx FVB/N;
 RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
 RA Strauberg R.;
 RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC019693; AAH19693.1; -;
 DR MGI; MGI:2384983; BC019693.
 DR InterPro; IPR008938; ARM.
 SQ SEQUENCE 349 AA; 40166 MW; 9763D0331AD0F515 CRC64;
 Query Match 63.4%; Score 26; DB 2; Length 349;
 Best Local Similarity 100.0%; Pred. No. 8.4e-20;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 KMKKNHMGPPMSILQEHIGVKKKEEL 27

```

Db          40 KMWKXHMGPFWMSIIQEHIGVKKKEEL 65
|||||
RESULT 4
ID 08CCT5 PRELIMINARY; PRT; 408 AA.
AC 08CCT5;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Mus musculus adult male olfactory brain cDNA, RIKEN full-length
DE enriched library, clone:6430400D6 product:hypothetical ARM repeat
DE structure containing protein, full insert sequence.
GN Name:BC019693
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Mech. Enzymol. 303:19-44 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RA The FANTOM Consortium;
RT "The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630 (2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Atzawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Kono H., Akiyama J., Nishi K., Kitagawa T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,
RA Yamamoto K., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771 (2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RA Adachi J., Atzawa K., Akiyama T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hitozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kaizuka T.,
RA Kato H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,

```

```

RA Nihi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai K., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK032134; BAC27721.1; -.
DR MGD; MGI:2384983; BC019693.
DR InterPro; IPR008938; ARM.
KW Hypothetical protein.
SQ SEQUENCE 408 AA; 4668 MM; 42AB9EB13CA1FE67 CRC64;
Query Match 63.4%; Score 26; DB 2; Length 408;
Best Local Similarity 100.0%; Pred. No. 9.6e-20;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 KMWKXHMGPFWMSIIQEHIGVKKKEEL 27
Db 99 KMWKXHMGPFWMSIIQEHIGVKKKEEL 124
|||||
RESULT 5
ID 096BS5 PRELIMINARY; PRT; 349 AA.
AC 096BS5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE FLN10359 protein.
GN Name:FLN10359;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=22388257; PubMed=12477932;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat M.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedon T.B., Toshitsuki S., Carninci P., Prange C.,
RA Rata S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Skolnick D., Green E.D., Schmutz J., Buterfield Y.S.,
RA Krywinski M.I., Skolnick D., Schmutz J., Myers R.M., Schein J.E.,
RA Jones S.J., Maira M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX STRAUBEERG R.;
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC011983; AAH11983.1; -.
DR InterPro; IPR008938; ARM.
SQ SEQUENCE 349 AA; 39921 MM; 3A359597FF7079EB CRC64;
Query Match 48.8%; Score 20; DB 2; Length 349;
Best Local Similarity 100.0%; Pred. No. 2.8e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 EKNWKXHMGPFWMSIIQEHIG 20
|||||

```

```

DB      39 EKNWKNHMGPFMSILOEHIG 58
RESULT 6
ID      Q8N7L7      PRELIMINARY;      PRT;      897 AA.
Q8N7L7
AC      Q8N7L7;
DT      01-OCT-2002 (TRENBLrel. 22, Created)
DT      01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT      01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE      Hypothetical protein FLJ40893.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Lymph;
RX      PubMed=14702039;
RA      Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA      Wakematsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA      Sekine M., Ohtsuka T., Nishii T., Shibahara T., Tanaka T., Ishii S.,
RA      Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA      Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA      Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA      Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA      Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA      Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA      Tanai H., Kimura M., Watanabe M., Hirooka S., Chiba Y., Ishida S.,
RA      Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hotsuta T., Kusano Y.,
RA      Taguchi K., Takahashi-Fujii A., Hara R., Tanase T., Nomura Y.,
RA      Toshiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA      Musshino K., Yuki H., Oshima A., Sasaki N., Aotsuka S.,
RA      Yoshioka Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
RA      Moriya S., Montiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA      Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA      Hishigaki H., Watanabe K., Sugiyama A., Takemoto M., Kawakami B.,
RA      Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA      Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA      Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,
RA      Kawabata A., Hikiji T., Kobatake N., Hiragaki H., Ikema Y., Okamoto S.,
RA      Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA      Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA      Togsashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA      Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA      Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
RA      Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,
RA      "Complete sequencing and characterization of 21,243 full-length human
RT      cDNAs."
RL      Nat. Genet. 36:40-45(2004).
DR      EMBL; AK098212; BAC05261.1; -.
DR      InterPro; IPR008938; ARM.
SQ      SEQUENCE 897 AA; 101574 MW; 5FF6A94FE8855995 CRC64;

Query Match      48.8%; Score 20; DB 2; Length 897;
Best Local Similarity 100.0%; Pred. No. 6,4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY      1 EKNWKNHMGPFMSILOEHIG 20
DB      587 EKNWKNHMGPFMSILOEHIG 606

RESULT 7
ID      Q6P197      PRELIMINARY;      PRT;      1106 AA.
Q6P197
AC      Q6P197;
DT      05-JUL-2004 (TRENBLrel. 27, Created)
DT      05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT      05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE      FLJ10359 protein (Fragment).
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Lymph;
RX      MEDLINE=22386257; PubMed=12477932;
RA      Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA      Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA      Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA      Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
RA      Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA      Stappleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA      Brownstein M.J., Ueidi T.B., Toohilyuk S., Carninci P., Prange C.,
RA      Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA      Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA      Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA      Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA      Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA      Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA      Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA      Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
RA      Kravtsovskiy M.I., Skalska U., Smallov D.E., Schermer A., Schein J.E.,
RA      Jones S.J., Marra M.A.;
RA      "Generation and initial analysis of more than 15,000 full-length human
RT      and mouse cDNA sequences."
RL      Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Lymph;
RX      Strausberg R.;
RL      Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC065205; AAH65205.1; -.
DR      InterPro; IPR008938; ARM.
FT      NON TER
SQ      SEQUENCE 1106 AA; 125359 MW; 09F7CE94042302C4 CRC64;

Query Match      48.8%; Score 20; DB 2; Length 1106;
Best Local Similarity 100.0%; Pred. No. 7,6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY      1 EKNWKNHMGPFMSILOEHIG 20
DB      796 EKNWKNHMGPFMSILOEHIG 815

RESULT 8
ID      AAH65205      PRELIMINARY;      PRT;      1106 AA.
AAH65205
AC      AAH65205;
DT      02-MAR-2004 (TRENBLrel. 27, Created)
DT      02-MAR-2004 (TRENBLrel. 27, Last sequence update)
DT      02-MAR-2004 (TRENBLrel. 27, Last annotation update)
DE      FLJ10359 protein (Fragment).
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Lymph;
RX      MEDLINE=22386257; PubMed=12477932;
RA      Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA      Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA      Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA      Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
RA      Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA      Stappleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA      Brownstein M.J., Ueidi T.B., Toohilyuk S., Carninci P., Prange C.,
RA      Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA      Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA      Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA      Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA      Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA      Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA      Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA      Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
RA      Kravtsovskiy M.I., Skalska U., Smallov D.E., Schermer A., Schein J.E.,
RA      Jones S.J., Marra M.A.;
RA      "Generation and initial analysis of more than 15,000 full-length human
RT      and mouse cDNA sequences."
RL      Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Lymph;
RX      Strausberg R.;
RL      Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC065205; AAH65205.1; -.
DR      InterPro; IPR008938; ARM.
FT      NON TER
SQ      SEQUENCE 1106 AA; 125359 MW; 09F7CE94042302C4 CRC64;

```

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Maitra M.A.,
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RA Strausberg R.,
 RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC065205; AA065205.1; -.
 FT NON TER 1
 SO SEQUENCE 1106 AA; 125359 MW; 09F7CE94042302C4 CRC64;
 Query Match 48.8%; Score 20; DB 2; Length 1106;
 Best Local Similarity 100.0%; Pred. No. 7,6e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EKNWKNHNGPMSIIQEHIG 20
 Db 796 EKNWKNHNGPMSIIQEHIG 815
 RESULT 9
 Q8BIJ2 PRELIMINARY; PRT; 163 AA.
 ID Q8BIJ2
 AC Q8BIJ2
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE Mus musculus B6-derived CD11+ve dendritic cells cDNA, RIKEN full-
 DE length enriched library, clone: F730048A22 product: weakly similar to
 DE BA93J16.3 (NOVEL KRAB BOX CONTAINING ZINC FINGER GENE).
 GN Name=BB114266;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=99279253; PubMed=10349636;
 RA Carninci P., Hayashizaki Y.,
 RT "High-efficiency full-length cDNA cloning";
 RL Mech. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=21085660; PubMed=11217851;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX The FANTOM Consortium,
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=20499374; PubMed=11042159;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes";

RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=20530913; PubMed=11076861;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama U., Nishi K., Katsunari T., Tashiro H., Itoh M.,
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwaigi K.,
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
 RT "RIKEN integrated sequence analysis (RISA) system-384 format
 RT sequencing pipeline with 384 multicapillary sequencer";
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hirooka T., Hirozane T.,
 RA Horii F., Imocani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Koyama S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK089532; BA040916.1; -.
 DR HSSP; P15822; IBBO.
 DR MGD; MGI:2143362; BB114266.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00096; zf_C2H2; 5.
 DR ProDom; PD000003; Znf_C2H2; 5.
 DR SMART; SM00355; Znf_C2H2; 5.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 5.
 SO SEQUENCE 163 AA; 19461 MW; A1ABFD2BCD74C044 CRC64;
 Query Match 19.5%; Score 8; DB 2; Length 163;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 29 SHOSQLTA 36
 Db 116 SHOSQLTA 123
 RESULT 10
 Q8BIJ2 PRELIMINARY; PRT; 354 AA.
 ID Q8BIJ2
 AC Q8BIJ2
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE Mus musculus ES cells cDNA, RIKEN full-length enriched library,
 DE clone: C330026E23 product: weakly similar to BA93J16.3 (NOVEL KRAB BOX
 DE CONTAINING ZINC FINGER GENE).
 GN Name=BB114266;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=99279253; PubMed=10349636;
 RA Carninci P., Hayashizaki Y.,

RT "High-efficiency full-length cDNA cloning."
 RL Mech. Enzymol. 303:19-44(1999).
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=2108560; PubMed=11217851;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RA The FANTOM Consortium,
 RT "The RIKEN Genome Exploration Research Group Phase I & II Team;
 RA "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=2049374; PubMed=11042159;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes."
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=20530913; PubMed=11076861;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama Y., Nishi K., Katsunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kasliwalgi K.,
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Wachihi K.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka K., Matsura S., Kawai U.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kita A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multichannel sequencer."
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kanakawa T.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M., Koya S.,
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK049344; BAC3696.1; -;
 DR HSSP; P08046; 1P21.
 DR MGD; MGI:2143362; B114266.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR00109; KRAB.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF01352; KRAB; 1.
 DR Pfam; PF00966; ZF-C2H2; 7.
 DR ProDom; PD000003; Znf_C2H2; 6.
 DR SMART; SM00349; KRAB; 1.
 DR SMART; SM00355; Znf_C2H2; 7.
 DR PROSITE; PS00805; KRAB; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 7.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 7.
 SO SEQUENCE 354 AA; 41447 MW; F2C93DEA421B43B4 CRC64;

Query Match 19.5%; Score 8; DB 2; Length 354;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 29 SHOSQTLA 36
 DB 307 SHOSQTLA 314
 RESULT 11
 Q7NSB2 PRELIMINARY; PRT; 445 AA.
 AC Q7NSB2;
 DT 01-MAR-2004 (TRENBLREL. 26, Created)
 DT 01-MAR-2004 (TRENBLREL. 26, Last sequence update)
 DE 01-MAR-2004 (TRENBLREL. 26, Last annotation update)
 DE Similar to quinol oxidase subunit I.
 GN OrderedLocustNames=piu2043;
 OS Photorhabdus luminescens (subsp. laumondii).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Photorhabdus.
 OX NCB1_TaxID=141679;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TT01;
 RX MEDLINE=22957627; PubMed=14528314;
 RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
 RA Taourit S., Bocs S., Bouraux-Eude C., Chandler M., Charles J.-F.,
 RA Dasse E., Derose R., Derzelle S., Freysinet G., Gaudriault S.,
 RA Medigue C., Lanois A., Powell K., Siguler F., Vincent R., Wingate V.,
 RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;
 RT "The genome sequence of the entomopathogenic bacterium Photorhabdus
 RT luminescens."
 RL Nat. Biotechnol. 21:1307-1313(2003).
 DR EMBL; BX571865; CAE14336.1; -;
 DR Photoblast; plu2043; -;
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR002585; Bac_Ubq_Cox.
 DR Pfam; PF01654; Bac_Ubq_Cox; 1.
 KW Complete proteome.
 SQ SEQUENCE 445 AA; 49978 MW; BDF82F209B973FA CRC64;
 Query Match 19.5%; Score 8; DB 2; Length 445;
 Best Local Similarity 100.0%; Pred. No. 3.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 34 LTAFFLEA 41
 DB 103 LTAFFLEA 110
 RESULT 12
 Q7NRX0 PRELIMINARY; PRT; 448 AA.
 AC Q7NRX0;
 DT 01-MAR-2004 (TRENBLREL. 26, Created)
 DT 01-MAR-2004 (TRENBLREL. 26, Last sequence update)
 DT 01-MAR-2004 (TRENBLREL. 26, Last annotation update)
 DE Cyanide insensitive terminal oxidase (EC 1.10.3.-).
 GN Name=cioA; OrderedLocustNames=CV3658;
 OS Chromobacterium violaceum.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Chromobacterium.
 OX NCB1_TaxID=536;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 12472 / DSM 30191;
 RX MEDLINE=22882880; PubMed=14500782;
 RA Vasconcelos A.T.R., de Almeida D.F., Hungria M., Guimaraes C.T.,
 RA Antonio R.V., Almeida F.C., de Almeida L.G.P., de Almeida R.,

RA Alves-Gomes J.A., Andrade E.M., Araripe J., de Araujo M.F.F.,
 RA Ascoli-Filho S., Azevedo V., Baptista A.J., Bataus L.A.M.,
 RA Barreira J.S., Belo A., van den Berg C., Bogio M., Bonato S.,
 RA Bordignon J., Brito D.M., Brito C.A., Brocchi M., Burity H.A.,
 RA Camargo A.A., Cardoso D.D.P., Carneiro N.P., Carraro D.M.,
 RA Carvalho C.M.B., Cascardo J.C.M., Cavada B.S., Chastre L.M.O.,
 RA Creczynski-Pasa T.B., Cunha-Junior N.C., Fagundes N., Falcao C.L.,
 RA Pantunatti F., Farías I.P., Felipe M.S., Ferrari L.P., Ferro J.A.,
 RA Ferro M.I.T., Franco G.R., Freitas N.S.A., Furlan L.R.,
 RA Gazzinelli R.T., Gomes E.A., Goncalves P.R., Grangeli T.B.,
 RA Grattapaglia D., Griebard E.C., Hanna E.S., Jardim S.N., Laurino J.,
 RA Leal L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.P.,
 RA Madeira H.M.F., Manfio G.P., Maranhao A.O., Martins W.S.,
 RA di Mauro S.M.Z., de Medeiros S.R.B., Melsen R.V., Moreira M.A.M.,
 RA Nascimento F.P., Nicolas M.F., Oliveira J.G., Oliveira S.C.,
 RA Palhao R.F.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,
 RA Pereira M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Potrich D.P.,
 RA Ramalho-Neto C.E., Reis A.M.M., Rigo L.U., Rondinelli E.,
 RA Santos E.B.P., Santos F.R., Schneider M.P.C., Senarez H.N.,
 RA Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,
 RA Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,
 RA Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Urmenyi T.,
 RA Vettore A., Wasson R., Zaha A., Simpson A.J.G.,
 RA "The complete genome sequence of *Chromobacterium violaceum* reveals
 RT remarkable and exploitable bacterial adaptability.",
 RL Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).
 DR EMBL: AE016923; AAC61320.1; -.
 DR GO: GO:0016020; C:membrane; IEA.
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.
 DR GO: GO:0006118; P:electron transport; IEA.
 DR InterPro: IPR002585; Bac_Ubp_Cox.
 DR Pfam: PF01654; Bac_Ubp_Cox; I.
 KW Complete proteome: Oxidoreductase.
 SQ SEQUENCE 448 AA; 49586 MW; 0803368EB45024D0 CRC64;

Query Match 19.5%; Score 8; DB 2; Length 448;
 Best Local Similarity 100.0%; Pred. No. 3.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 LTAPFLA 41
 Db 104 LTAPFLA 111

RESULT 13
 Q98DW1 PRELIMINARY; PRT; 453 AA.
 AC Q98DW1.
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Quinol oxidase subunit I.
 GN OrderedLocusNames=mlr4524;
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 CC Phyllobacteriaceae; Mesorhizobium.
 CX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082936; PubMed=11214974;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti (supplement).";
 RL DNA Res. 7:381-406(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;

RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti.",
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP003004; BAB51159.1; -.
 DR GO: GO:0016020; C:membrane; IEA.
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.
 DR GO: GO:0006118; P:electron transport; IEA.
 DR InterPro: IPR002585; Bac_Ubp_Cox.
 DR Pfam: PF01654; Bac_Ubp_Cox; I.
 KW Complete proteome.
 SQ SEQUENCE 453 AA; 49427 MW; 1D2439BE634BEA0F CRC64;

Query Match 19.5%; Score 8; DB 2; Length 453;
 Best Local Similarity 100.0%; Pred. No. 3.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 LTAPFLA 41
 Db 100 LTAPFLA 107

RESULT 14
 Q87H28 PRELIMINARY; PRT; 454 AA.
 AC Q87H28.
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Cytochrome BD2, subunit I.
 GN OrderedLocusNames=VPAL137;
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 CC Vibrionaceae; Vibrrio.
 CX NCBI_TaxID=670;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIMD 2210633 / Serotype O3:K6;
 RX MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Nishimura M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.,
 RT "Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism
 RT distinct from that of *V. cholerae*.",
 RL Lancet 361:743-749(2003).
 DR EMBL: AP005088; BAC62480.1; -.
 DR GO: GO:0016020; C:membrane; IEA.
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.
 DR GO: GO:0006118; P:electron transport; IEA.
 DR InterPro: IPR002585; Bac_Ubp_Cox.
 DR Pfam: PF01654; Bac_Ubp_Cox; I.
 KW Complete proteome.
 SQ SEQUENCE 454 AA; 50975 MW; 318223526B9A946 CRC64;

Query Match 19.5%; Score 8; DB 2; Length 454;
 Best Local Similarity 100.0%; Pred. No. 3.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 LTAPFLA 41
 Db 103 LTAPFLA 110

RESULT 15
 Q7WB10 PRELIMINARY; PRT; 462 AA.
 AC Q7WB10.
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

```

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cytochrome oxidase subunit I.
GN Name-cioA; Synonyms-gxvA; OrderedLocusNames-BPP1024;
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=12822 / ATCC BAA-587;
RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leathem S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrett B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL: BX640426; CAE36324.1; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0006118; P:electron transport; IEA.
DR InterPro: IPR002585; Bac_Ubq_Cox.
DR Pfam: PF01654; Bac_Ubq_Cox; 1.
KW Complete proteome.
SQ SEQUENCE 462 AA; 49902 MW; 631FAC87419163ED CRC64;

Query Match 19.5%; Score 8; DB 2; Length 462;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 34 LTAFLEA 41
DB 103 LTAFLEA 110

```

Search completed: February 1, 2005, 15:31:02
 Job time : 71.5556 secs

CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention

CC
 CC Sequence 349 AA;
 CC
 CC Query Match 100.0%; Score 41; DB 4; Length 349;
 CC Best Local Similarity 100.0%; Pred. No. 1.6e-35;
 CC Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NCLYKIFLPTQHFISKERAEALMPLVDQLENRLGSEK 41
 Db 202 NCLYKIFLPTQHFISKERAEALMPLVDQLENRLGSEK 242

RESULT 2
 ADE08012
 ID ADE08012 standard; protein; 1149 AA.
 AC ADE08012;
 AC
 AC ADE08012;
 AC
 DT 29-JAN-2004 (first entry)
 DT
 DE Novel protein (useful for identifying genetic disorders) #167.
 DE
 KW novel gene; novel protein; tissue marker; molecular weight marker;
 KW chromosome marker; genetic disorder.
 OS Unidentified.
 OS
 PN WO2003054152-A2.
 PN
 PD 03-JUL-2003.
 PD
 PP 10-DEC-2002; 2002WO-US039555.
 PP
 PR 10-DEC-2001; 2001US-0339739P.
 PR 11-DEC-2001; 2001US-0339453P.
 PR 14-MAR-2002; 2002US-0365091P.
 PR 14-MAR-2002; 2002US-0365384P.
 PR 12-APR-2002; 2002US-0372381P.
 PR 12-APR-2002; 2002US-0372615P.
 PR 22-APR-2002; 2002US-00128558.
 PR 24-APR-2002; 2002US-0376045P.
 PR
 PA (HYSE-) HYSEQ INC.
 PA
 PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
 PI Ghosh M, Xue AJ, Wehrman T, Wang G, Zhou P, Drmanac RT, Wang Z;
 PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ,
 PI
 XX WPI: 2003-569235/53.
 XX N-PSDB; ADE07101.
 XX
 DR New polynucleotides, useful for expressing recombinant proteins for
 DR PT analysis, characterization or therapeutic use, or as markers for tissues
 DR PT in which the corresponding protein is preferentially expressed.
 PT
 XX Claim 20; SEQ ID NO 1078; 1177bp; English.
 XX
 CC The invention comprises the amino acid and coding sequences of novel
 CC proteins. The DNA and protein sequences of the invention are useful as:

CC markers for tissues in which the corresponding protein is preferentially
 CC expressed; as molecular weight markers on gels; as chromosome markers or
 CC tags; to identify chromosomes or to map related gene positions; and to
 CC compare with endogenous DNA sequences in patients to identify potential
 CC genetic disorders. The present amino acid sequence represents a protein
 CC of the invention.

CC
 CC Sequence 1149 AA;
 CC
 CC Query Match 100.0%; Score 41; DB 7; Length 1149;
 CC Best Local Similarity 100.0%; Pred. No. 4.6e-35;
 CC Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NCLYKIFLPTQHFISKERAEALMPLVDQLENRLGSEK 41
 Db 1002 NCLYKIFLPTQHFISKERAEALMPLVDQLENRLGSEK 1042

RESULT 3
 AAM54099
 ID AAM54099 standard; protein; 515 AA.
 AC AAM54099;
 AC
 AC AAM54099;
 AC
 DT 28-SEP-1998 (first entry)
 DT
 DE Homo sapiens BAP28 sequence.
 DE
 KW BARD1, ring protein; BRCA1, breast cancer; risk; diagnosis.
 KW
 OS Homo sapiens.
 OS
 PN WO9812327-A2
 PN
 PD 26-MAR-1998.
 PD
 PP 19-SEP-1997; 97WO-US016842.
 PP
 PR 20-SEP-1996; 96US-0025296P.
 PR 03-APR-1997; 97US-0042611P.
 PR 04-APR-1997; 97US-0042985P.
 PR
 PA (TEXA) UNIV TEXAS SYSTEM.
 PA
 PI Bowcock AM, Baer R;
 PI
 XX WPI: 1998-230317/20.
 XX N-PSDB; AAV24135.
 XX
 DR DNA sequence encoding BARD1, B123, BE2, BE14, BE31 or BE445 - which as
 DR PT breast cancer antigen, BRCA1, binding proteins are useful to identify
 DR PT patient having or at risk of developing cancer.
 PT
 XX Disclosure; Page 287-288; 348pp; English.
 XX
 CC The sequence is that of a protein which can be used in the preparation of
 CC the recombinant breast cancer antigen, BRCA1, binding proteins BARD1,
 CC B123, BE2, BE14, BE31 or BE445, or a composition for the detection of a
 CC BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid sequence, specifically
 CC a wild type BARD1 composition for the detection or purification of BRCA1,
 CC useful to identify a patient having, or at risk of developing cancer.
 CC BARD1 can be used in the preparation of an anti-BARD1 antibody, and in
 CC the detection and purification of a BRCA1 protein. BARD1, B123, BE2,
 CC BE14, BE31 or BE445 can be used in the identification of a binding protein
 CC agonist or antagonist that alters the binding of BARD1, B123, BE2, BE14,
 CC BE31 or BE445 to BRCA1 or the biological activity of the BRCA1-BARD1,
 CC B123, BE2, BE14, BE31 or BE445 complex. The antibodies can be used to
 CC detect BARD1, B123, BE2, BE14, BE31 or BE445, a specific anti-BARD1
 CC antibody can be used to identify a patient having or at risk of
 CC developing cancer

CC
 CC Sequence 515 AA;
 CC

Query Match 48.8%; Score 20; DB 2; Length 515;
Best Local Similarity 100.0%; Pred. No. 7.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NCLYKIFLPTQHFISKERA 20
Db 368 NCLYKIFLPTQHFISKERA 387

RESULT 4
AAB85029 standard; protein; 2144 AA.
XX ID AAB85029 standard; protein; 2144 AA.
XX AC AAB85029;
XX DT 06-AUG-2001 (first entry)
XX DE Protein encoded by BAP28 cDNA consisting of exons 1 to 45.
XX KW BAP28; prostate; tumour; cancer; diagnostic; genetic analysis.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Misc-difference 1694
FT Misc-difference 1854 /label= Ser or Asn
FT Misc-difference 1967 /label= Ala or Val
FT Misc-difference 2017 /label= Asp or Asn
FT Misc-difference 2017 /label= Gly or Glu
XX PN W0200100669-A2.
XX PD 04-JAN-2001.
XX PF 23-JUN-2000; 2000WO-IB001183.
XX PR 25-JUN-1999; 99US-0141323P.
XX PR 18-JAN-2000; 2000US-0176880P.
XX PA (GEST) GENSET.
XX PI Barry C, Bougueleret L, Chumakov I, Cohen-Akenine A;
XX DR WPI; 2001-367032/38.
XX DR N-PSDB; AAF83909, AAF83910.
XX PT New BAP28 polynucleotides and polypeptides overexpressed in prostate
PT cancer cells for diagnosing prostate tumours, e.g. by hybridization or
PT polymerase chain reaction assays.
XX PS Claim 14; Page 297-304; 349pp; English.
XX CC The invention is directed to BAP28 polypeptides, BAP28 polynucleotide
CC sequences and regulatory region located at the 3' and 5' ends of the
CC BAP28 coding region. The BAP28 polypeptides can be expressed by standard
CC recombinant methodology. BAP28 polynucleotides and polypeptides have been
CC found to be over expressed in prostate tumour cells, therefore levels of
CC BAP28 expression and/or activity may be assayed (e.g. by polymerase chain
CC reaction (PCR)) to diagnose patient suffering from or susceptible to
CC prostate cancer. Antibodies specific for the BAP28 polypeptides are
CC useful as diagnostic reagents. Allelic markers of the BAP28 gene are
CC useful in genetic analysis. The present sequence represents a protein
CC encoded by a first cDNA sequence of the BAP28 gene consisting of the
CC exons 1 to 45
XX SQ Sequence 2144 AA;

Query Match 48.8%; Score 20; DB 4; Length 2144;
Best Local Similarity 100.0%; Pred. No. 2.7e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NCLYKIFLPTQHFISKERA 20
Db 1997 NCLYKIFLPTQHFISKERA 2016

RESULT 5
ABU21620 standard; protein; 378 AA.
XX ID ABU21620 standard; protein; 378 AA.
XX AC ABU21620;
XX DT 19-JUN-2003 (first entry)
XX DE Protein encoded by Prokaryotic essential gene #7147.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Burkholderia fungorum.
XX PN W0200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX DR WPI; 2003-029926/02.
XX DR N-PSDB; ACA25490.
XX PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 25; SEQ ID NO 49544; 1766pp; English.
XX CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense-nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

XX ABB63215;
AC 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster polypeptide SEQ ID NO 16437.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.
XX
OS Drosophila melanogaster.
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PP 23-MAR-2001; 2001WO-US0009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
DR WPI; 2001-656860/75.
DR N-PSDB; ABL07318.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 16437; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC sequences (ABU01840-ABU16175) and the encoded proteins (ABU57737-
CC ABU72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1150 AA;
Query Match 17.1%; Score 7; DB 4; Length 1150;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 17 KERAAL 23
Db 315 KERAAL 321
RESULT 9
ABU15816
ID ABB15816 standard; protein; 32 AA.
XX
AC ABB15816;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polypeptide SEQ ID NO 4473.
XX
DE Human; nociceptive; cytoskeletal; dermatological; virologic;
KM immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnerary;
KM antiparkinsonian; antischistosomal; antineoplastic; antirheumatic; cancer;
KM antineuritic; hepatocellular; cerebroprotective; anti-inflammatory;
KM antidiabetic; antitumor; anticonvulsant; antifungal;
KM antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KM neurological disease; infection; neuroprotective; gene therapy; vaccine.

XX Homo sapiens.
OS
XX WO200159063-A2.
PN
XX 16-AUG-2001.
PD
XX
PF 17-JAN-2001; 2001WO-US001334.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225233P.
PR 14-AUG-2000; 2000US-0225244P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231245P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0232081P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234222P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.

PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239335P.
 PR 13-OCT-2000; 2000US-0239337P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0244747P.
 PR 08-NOV-2000; 2000US-0244752P.
 PR 08-NOV-2000; 2000US-0244761P.
 PR 08-NOV-2000; 2000US-0244776P.
 PR 08-NOV-2000; 2000US-0244786P.
 PR 08-NOV-2000; 2000US-0244523P.
 PR 08-NOV-2000; 2000US-0244524P.
 PR 08-NOV-2000; 2000US-0245252P.
 PR 08-NOV-2000; 2000US-0245253P.
 PR 08-NOV-2000; 2000US-0245272P.
 PR 08-NOV-2000; 2000US-0245282P.
 PR 08-NOV-2000; 2000US-0245332P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249246P.
 PR 17-NOV-2000; 2000US-0249255P.
 PR 17-NOV-2000; 2000US-0249257P.
 PR 17-NOV-2000; 2000US-0249259P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 01-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251859P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX

PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-541565/60.
 DR N-PSDB; ABA12142.
 XX
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system cancers
 PT and metacases.
 XX
 PS Claim 11; SEQ ID NO 4473; 1701pp + Sequence listing; English.
 XX
 CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
 CC (ABA1678-ABA18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
 CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
 CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections. Note: The sequence data for this patent did not form part of
 CC the pinned specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 32 AA;
 XX
 QY 4 YKIFLF 9
 Db 21 YKIFLF 26
 XX
 RESULT 10
 ABB40457
 ID ABB40457 standard; peptide; 36 AA.
 XX
 AC ABB40457;
 XX
 DT 04-FEB-2002 (first entry)
 XX
 DE Peptide #7963 encoded by human foetal liver single exon probe.
 XX
 KM Human, foetal liver; gene expression; single exon nucleic acid probe.
 XX
 OS Homo sapiens.
 XX
 PN W0200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000669.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DK;

XX WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human fetal liver.
XX
PS Claim 27; SEQ ID NO 33092; 639bp + Sequence Listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human fetal liver. The
CC present sequence is a peptide encoded by a single exon nucleic acid probe
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_ptc_sequences
XX
SQ Sequence 36 AA;
XX
Query Match 14.6%; Score 6; DB 4; Length 36;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 36 GGEKXF 41
DB 4 GGEKXF 9
XX
RESULT 11
AAM34177
ID AAM34177 standard; protein; 36 AA.
XX
AC AAM34177;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #8214 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KM genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000663.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
PI WPI; 2001-488897/53.
XX
DR
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human placenta.
XX
PS Claim 27; SEQ ID NO 34446; 654bp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENPs;
CC see AAI3315-AA157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of

CC human genetic disorders
XX
SQ Sequence 36 AA;
XX
Query Match 14.6%; Score 6; DB 4; Length 36;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 36 GGEKXF 41
DB 4 GGEKXF 9
XX
RESULT 12
AAM74002
ID AAM74002 standard; protein; 36 AA.
XX
AC AAM74002;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 34308.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000668.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
PI WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
PS Example 4; SEQ ID NO 34308; 658bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention
XX
SQ Sequence 36 AA;
XX
Query Match 14.6%; Score 6; DB 4; Length 36;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 36 GGEKXF 41
DB 4 GGEKXF 9
XX
RESULT 13
AAM61279

ID AAM61279 standard; protein; 36 AA.
XX
AC AAM61279;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 33384.
XX
KW Human; brain expressed exon; gene expression analysis; probe; microarray;
XX Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000667.
XX
PR 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX
PS Example 4; SEQ ID NO 33384; 650bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancer. The present sequence is a protein encoded by one of
XX the probes of the invention
XX
SQ Sequence 36 AA;
XX
Query Match 14.6%; Score 6; DB 4; Length 36;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 36 GGEKRF 41
DB 4 GGEKRF 9
XX
RESULT 14
ABG55759
ID ABG55759 standard; peptide; 36 AA.
XX
AC ABG55759;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver peptide, SEQ ID NO 34407.
XX
KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX hypercholesterolaemia; coronary heart disease.
XX
OS Homo sapiens.
XX
PN WO200157273-A2.

XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000664.
XX
PR 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488898/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human adult liver.
XX
PS Claim 27; SEQ ID NO 34407; 658bp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX measuring human gene expression in a sample derived from human adult
XX liver, comprising one of 13109 defined nucleotide sequences given in the
XX specification (or complements/fragments). The probe hybridises at high
XX stringency to a nucleic acid molecule expressed in the human adult liver.
XX (I) may be used for predicting, measuring and displaying gene expression
XX in samples derived from human adult liver. The genes identified may be
XX involved in genetic liver diseases such as cirrhosis,
XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
XX associated with coronary heart disease. ABG47348-ABG59930 represent human
XX liver single exon encoded peptides of the invention. Note: The sequence
XX information for this patent does not appear in the printed specification
XX but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 36 AA;
XX
Query Match 14.6%; Score 6; DB 4; Length 36;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 36 GGEKRF 41
DB 4 GGEKRF 9
XX
RESULT 15
ABG43897
ID ABG43897 standard; peptide; 36 AA.
XX
AC ABG43897;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human peptide encoded by genome-derived single exon probe SEQ ID 33562.
XX
KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hereditary spherocytosis; sarcoidosis; pulmonary haemosiderosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease.
XX
OS Homo sapiens.
XX
PN WO200186003-A2.

Search completed: February 1, 2005, 15:20:32
Job time : 71.7778 secs

```

XX 15-NOV-2001.
PD
XX
PF 30-JAN-2001; 2001MO-US000665.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples.
XX
PS Claim 27; SEQ ID NO 33562; 634bp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of probes
CC; the novel set of probes which hybridize at high stringency to a nucleic
CC acid expressed in the human lung; measuring gene expression in a sample
CC derived from human lung; comprising (a) contacting the array with a
CC collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of the
CC array; identifying exons in a eukaryotic genome, comprising (a)
CC algorithmically predicting at least one exon from genomic sequences of
CC the eukaryote; and (b) detecting specific hybridization of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridization to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene expression
CC analysis, and for identifying exons in a gene, particularly using human
CC lung derived mRNA and for the study of lung diseases such as asthma, lung
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
CC present sequence is a peptide/protein encoded by a single exon probe of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 36 AA;
Query Match 14.6%; Score 6; DB 5; Length 36;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 36 GGEKRF 41
Db 4 GGEKRF 9

```

This Page Blank (uspto)

Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 18 PCIAQFS 24
Db 597 PCIAQFS 603

RESULT 3

B96552
Unknown protein, 88255-88575 [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C/Accession: B96552
R/Theologian: A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 815-820, 2000
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A/Authors: Salberg, S.L.; Schwartz, J.R.; Shum, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A/Reference number: A6141; MUID:21016719; PMID:11130712
A/Status: Preliminary
A/Accession: B96552
A/Molecule type: DNA
A/Residues: 1-106 <STO>
A/Cross-references: UNIPROT:Q9SYE2; GB:AE005173; NID:g10092351; PIDN:AA012760.1; GSPDB:C
C/Genetics:
A/Gene: F5D21.10
A/Map position: 1

Query Match 14.6%; Score 6; DB 2; Length 106;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 VAMADD 30
Db 73 VAMADD 78

RESULT 4

S29883
Ribosomal protein L24 - Micrococcus luteus
C/Species: Micrococcus luteus, Micrococcus lysodeikticus
C/Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C/Accession: S29883
R/Ohama, T.; Muto, A.; Ogawa, S.
J. Mol. Evol. 29, 381-395, 1989
A/Title: Spectinomycin operon of Micrococcus luteus: evolutionary implications of organ
A/Reference number: S17070; MUID:90133967; PMID:2533272
A/Accession: S29883
A/Molecule type: DNA
A/Residues: 1-113 <OHM>
A/Cross-references: UNIPROT:P33103; EMBL:X17524; NID:g44425; PIDN:CAA35559.1; PID:g44427
C/Suprafamily: Escherichia coli ribosomal protein L24
C/Keywords: protein biosynthesis; ribosome

Query Match 14.6%; Score 6; DB 2; Length 113;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 RVTKHL 16
Db 41 RVTKHL 46

RESULT 5

BRAD4
early E3B 15.3K protein - human adenovirus 7 (strain Gomen)
C/Species: Mastadenovirus h7 (human adenovirus 7)

A/Note: host Homo sapiens (man)
C/Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 09-Jul-2004
C/Accession: E31830
R/Hong, J.S.; Mullis, K.G.; Engler, J.A.
Virology 167, 545-553, 1988
A/Title: Characterization of the early region 3 and fiber genes of Ad7.
A/Reference number: A94386; MUID:99073758; PMID:2849239
A/Accession: E31830
A/Molecule type: DNA
A/Residues: 1-135 <HON>

A/Cross-references: UNIPROT:P15135; GB:M23696; NID:g341012; PIDN:AA53253.1; PID:g576460
C/Suprafamily: adenovirus early E3B 14K protein
C/Keywords: early protein

Query Match 14.6%; Score 6; DB 1; Length 135;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QERVTK 14
Db 35 QERVTK 40

RESULT 6

E70385
conserved hypothetical protein aq_987 - Aquifex aeolicus
C/Species: Aquifex aeolicus
C/Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C/Accession: E70385
R/Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
V.
Nature 392, 353-358, 1998
A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A/Reference number: A70300; MUID:98196666; PMID:9537320
A/Accession: E70385
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-165 <AQF>
A/Cross-references: UNIPROT:O67112; GB:AE000717; NID:g2983492; PIDN:AA07079.1; PID:g2983
A/Experimental source: strain VF5
C/Genetics:
A/Gene: aq_987
C/Suprafamily: Pyrococcus abyssi hypothetical protein PAB0445

Query Match 14.6%; Score 6; DB 2; Length 165;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RLGGE 6
Db 38 RLGGE 43

RESULT 7

AG2091
glucosyltransferase [imported] - Nostoc sp. (strain PCC 7120)
C/Species: Nostoc sp. PCC 7120
A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C/Accession: AG2091
R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi,
Nakazaki, N.; Shimp, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anab
A/Reference number: AB1807; MUID:21595285; PMID:11759840
A/Accession: AG2091
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-169 <KUR>
A/Cross-references: UNIPROT:Q8YUP9; GB:BA000019; PIDN:BA873985.1; PID:g17131378; GSPDB:G
A/Experimental source: strain PCC 7120
C/Genetics:
A/Gene: all2286

C:Superfamily: Streptococcus pneumoniae beta-1,4-galactosyltransferase cps14G

Query Match 14.6%; Score 6; DB 2; Length 169;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 EKKFQE 10
|||||
Db 55 EKKFQE 60

RESULT 8

AE1538

transcription regulator TsrA/Acr family homolog lin0845 [imported] - *Listeria innocua*

C:Species: *Listeria innocua*
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004

C:Accession: AE1538

R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entlian, K.D.; Fehli, H.

Science 294, 849-852, 2001

A:Authors: Krefit, U.; Kunz, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tietze, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A.; Title: Comparative genomics of *Listeria* species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AE1538

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-184 <GUA>

A:Cross-references: UNIPROT:Q92DQ8; GB:AL592022; PIDN:CAC96077.1; PID:G16413296; GSPDB:C

A:Experimental source: strain Clp11262

C:Genetics:

A:Gene: lin0845

Query Match 14.6%; Score 6; DB 2; Length 184;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LGGEK 7
|||||
Db 73 LGGEK 78

RESULT 9
F87295
conserved hypothetical protein CC0375 [imported] - *Caulobacter crescentus*

C:Species: *Caulobacter crescentus*
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004

C:Accession: F87295

R:Niernann, W.C.; Feldblym, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.B.; Lamb, M.T.; Deboy, R.T.; Dodson, R.U.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of *Caulobacter crescentus*.

A:Reference number: AB7249; MUID:21173698; PMID:11259647

A:Accession: F87295

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-204 <STO>

A:Cross-references: UNIPROT:Q9AB57; GB:AB005673; NID:G13421530; PIDN:AK23262.1; GSPDB:C

C:Genetics:

A:Gene: CC0375

Query Match 14.6%; Score 6; DB 2; Length 204;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RLGGEE 6
|||||

Db 181 RLGGEE 186

RESULT 10

AF0199
probable lipoprotein YP01635 [imported] - *Yersinia pestis* (strain CO92)

C:Species: *Yersinia pestis*
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004

C:Accession: AF0199

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, N.

Nature 413, 523-527, 2001

A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AF0199

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-207 <KIR>

A:Cross-references: UNIPROT:Q8ZFQ8; GB:AL590842; PIDN:CAC90457.1; PID:G15979673; GSPDB:C

C:Genetics:

A:Gene: YP01635

Query Match 14.6%; Score 6; DB 2; Length 207;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 SVAMAD 29
|||||
Db 16 SVAMAD 21

RESULT 11
AG1173
hypothetical protein lmo0791 [imported] - *Listeria monocytogenes* (strain EGD-e)

C:Species: *Listeria monocytogenes*
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004

C:Accession: AG1173

R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entlian, K.D.; Fehli, H.

Science 294, 849-852, 2001

A:Authors: Krefit, U.; Kunz, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tietze, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A.; Title: Comparative genomics of *Listeria* species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AG1173

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-216 <GUA>

A:Cross-references: UNIPROT:Q8Y8V1; GB:NC_003210; PIDN:CAC98869.1; PID:G16410180; GSPDB:C

C:Genetics:

A:Gene: lmo0791

Query Match 14.6%; Score 6; DB 2; Length 216;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 29 DDSLWK 34
|||||
Db 191 DDSLWK 196

RESULT 12
AB7990
protein W05H12.1 [imported] - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004

C:Accession: AB7990

R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode *C. elegans*: a platform for investigating biology.
A:Reference number: A75000; MUID:99069613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.elegans/
A:Accession: AB7990

A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-217 <STO>
A/Cross-references: UNIPROT:O62394; GB:chtr_I; PIDN:CAB04920.1; PID:G3880544; GSPDB:GN000
C/Genetics:
A/Gene: W05H12.1
A/Map position: 1

Query Match 14.6%; Score 6; DB 2; Length 217;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 IAFPSV 25
17 IAFPSV 22

RESULT 13
AH1530
hypothetical protein lin0784 [imported] - *Listeria innocua* (strain Clijp1262)
C/Species: *Listeria innocua*
C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C/Accession: AH1530
R/Glauber, P.; Frangoul, L.; Buchtriser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, R.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fehl, H.; Jones, L.M.; Karet, U.
Science 294, 849-852, 2001
A/Authors: Krett, J.; Kuhn, M.; Kunst, F.; Kurapkac, G.; Madueno, E.; Maitournam, A.; Mok, C.; Schlueter, T.; Simoes, N.; Tietzer, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A./Title: Comparative genomics of *Listeria* species.
A/Reference number: AB1077; MUID:21537279; PMID:11679669
A/Accession: AH1530
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-218 <GUA>
A/Cross-references: UNIPROT:Q92DM9; GB:AL592022; PIDN:CAC96016.1; PID:G16413235; GSPDB:G
A/Experimental source: strain Clijp1262
C/Genetics:
A/Gene: lin0784

Query Match 14.6%; Score 6; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 DDSLWK 34
192 DDSLWK 197

RESULT 14
B81448
moeb/thif family protein Cj0294 [imported] - *Campylobacter jejuni* (strain NCTC 11168)
C/Species: *Campylobacter jejuni*
C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C/Accession: B81448
R/Parikh, J.; Wren, B.W.; Mungall, K.; Kettle, J.M.; Churcher, C.; Basham, D.; Chilling, C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrett, Nature 403, 665-668, 2000
A/Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyf
A/Reference number: AB1250; MUID:20150912; PMID:10688204
A/Accession: B81448
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-219 <PAR>
A/Cross-references: UNIPROT:Q9PIK5; GB:AL139074; GB:AL111168; NID:G6967505; PIDN:CAB7276
A/Experimental source: serotype O2, strain NCTC 11168
C/Genetics:
A/Gene: Cj0294

Query Match 14.6%; Score 6; DB 2; Length 219;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GEEKFQ 9
14 GEEKFQ 19

RESULT 15
AC3293
thiamine-phosphate diphosphorylase (EC 2.5.1.3) [imported] - *Brucella melitensis* (strain C/Species: *Brucella melitensis*
C/Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C/Accession: AC3293
R/DelVecchio, V.G.; Kaput, V.; Redkar, R.J.; Patra, G.; Mujeer, C.; Los, T.; Ivanova, T.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Lelessec, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A/Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
A/Reference number: AD3252; PMID:11756688
A/Accession: AC3293
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-221 <KUR>
A/Cross-references: UNIPROT:Q8Y1W0; UNIPROT:Q8FYZ8; GB:AE008917; PIDN:AAL51510.1; PID:G1
C/Experimental source: strain 16M
C/Genetics:
A/Gene: BMEI0329
A/Map position: 1
C/Keywords: transferase

Query Match 14.6%; Score 6; DB 2; Length 221;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 IPCIAQ 22
165 IPCIAQ 170

Search completed: February 1, 2005, 15:33:08
Job time : 15.5556 secs